Law, Governance and Technology Series 51
Issues in Privacy and Data Protection

Paweł Księżak
Sylwia Wojtczak

Toward a Conceptual Network for the Private Law of Artificial Intelligence

Springer

Law, Governance and Technology Series

Issues in Privacy and Data Protection

Volume 51

Issues in Privacy and Data Protection aims at publishing peer reviewed scientific manuscripts that focus upon issues that engage into an analysis or reflexion related to the consequences of scientific and technological developments upon the private sphere, the personal autonomy and the self-construction of humans with data protection and privacy as anchor points. The objective is to publish both disciplinary, multidisciplinary and interdisciplinary works on questions that relate to experiences and phenomena that can or could be covered by legal concepts stemming from the law regarding the protection of privacy and/or the processing of personal data. Since both the development of science and technology, and in particular information technology (ambient intelligence, robotics, artificial intelligence, knowledge discovery, data mining, surveillance, etc.), and the law on privacy and data protection are in constant frenetic mood of change (as is clear from the many legal conflicts and reforms at hand), we have the ambition to reassemble a series of highly contemporary and forward-looking books, wherein cutting edge issues are analytically, conceptually and prospectively presented.

Paweł Księżak • Sylwia Wojtczak

Toward a Conceptual Network for the Private Law of Artificial Intelligence

Springer

Paweł Księżak
Civil Law Department
University of Lodz
Łódź, Poland

Sylwia Wojtczak
Department of Legal Policy
University of Lodz
Łódź, Poland

This work was supported by National Science Centre (Poland) according to the agreement UMO-2018/29/B/HS5/00421.

ISSN 2352-1902 ISSN 2352-1910 (electronic)
Law, Governance and Technology Series
ISSN 2352-1929 ISSN 2352-1937 (electronic)
Issues in Privacy and Data Protection
ISBN 978-3-031-19446-7 ISBN 978-3-031-19447-4 (eBook)
https://doi.org/10.1007/978-3-031-19447-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

1 Introduction: Is a New Conceptual Network Necessary to Adapt the Civil (Private) Law to the Development of AI and Robotics Development? 1
References 10
Books and Articles 10
Documents 11

2 Artificial Intelligence and Legal Subjectivity 13
2.1 Introduction 13
2.2 Alleged Hierarchy: (AI?), Human Beings, (AI?), Juristic Persons, (AI?), Animals 17
2.3 Sentience and Reason 19
2.4 Presence/Participation in Social Life 23
2.5 Legal Subjectivity as a Social Fact 25
2.6 Does AI Participate or Is Present in Social Life? 28
2.7 Should AI Be Endowed with Legal Subjectivity? 30
2.8 What Form of Legal Subjectivity Should AI Have? Electronic Persons, Synthetic Persons Etc. 32
References 33
Books and Articles 33
Documents 35
Online Sources 35

3 Will and Discernment 37
3.1 Introduction 37
3.2 Free Will and Discernment of AI? 41
3.2.1 Free Will 41
3.2.2 Discernment 49

References . . . 50
Books and Articles . . . 50
Documents . . . 51
Online Sources . . . 52

4 Capacity for Juridical Acts . . . 53
4.1 Introduction . . . 53
4.2 Capacity for Juridical Acts of AI: Theoretical and Legal Bases . . . 56
4.3 A Legal Capacity of AI and Its Capacity for Juridical Acts as a Function of Registration . . . 63
4.4 Capacity for Juridical Acts by a Human User of AI . . . 67
4.5 Capacity for Juridical Acts by the Juridical Person Using the AI . . . 72
References . . . 74
Books and Articles . . . 74
Documents . . . 75

5 Consent . . . 77
5.1 Introduction . . . 77
5.2 Attributability . . . 79
5.2.1 The Construct . . . 79
5.2.2 Exceeding the Scope of Authorization . . . 85
5.2.3 Acting Outside Registration . . . 87
5.2.4 Acting as a Legal Person's Body . . . 88
5.2.5 AI-Representative Acting in Its Own Name . . . 90
5.3 AI's Intent and Declaration of Intent . . . 91
5.4 Contracts A2A (AI-to-AI) . . . 95
5.5 Defects in the Declaration of Intent. Vitiated Consent or Intention . . . 95
5.5.1 The Concept . . . 95
5.5.2 Mistake and Fraud . . . 98
5.5.3 Threats . . . 102
5.5.4 Unfair Exploitation . . . 103
References . . . 106
Books and Articles . . . 106
Documents . . . 107

6 Personal Interests of AI . . . 109
6.1 Introduction . . . 109
6.2 The Possible Types of Personal Interests of AI . . . 115
6.2.1 Existence and Procreation . . . 115
6.2.2 Personal Interests Related to Consciousness, Emotions and Embodiment . . . 119
6.2.3 Personal Interests Implied by Social Relations: Identity and Reputation . . . 127
6.3 Personal Interests of AI After Its "Death" . . . 128

References . . . 129
Books and Articles . . . 129
Documents . . . 130

7 Copyright . . . 131
7.1 Introduction . . . 131
7.2 The Work: The Founding Category of Copyright . . . 134
7.3 AI and the Work. Existing Concepts . . . 140
7.4 AI as an Author . . . 142
7.5 The Proposal . . . 143
References . . . 147
Books and Articles . . . 147
Documents . . . 149

8 Property . . . 151
8.1 Introduction . . . 151
8.2 AI as an Owner of Property . . . 152
8.3 AI as Property . . . 157
8.4 The Will of AI Versus the Will of the Owner . . . 160
8.5 AI as a Subject of Joint Ownership . . . 169
8.5.1 Shared AI in the Household . . . 170
8.5.2 Joint AI in a Company . . . 177
8.6 AI as a Household and Family Member . . . 179
8.7 AI as a Possessor of Property . . . 181
8.8 AI as an Owner of AI . . . 183
8.9 Succession of Rights of AI . . . 183
8.10 The User's Death and the Succession of the Personalized AI . . . 185
References . . . 186
Books and Articles . . . 186
Documents . . . 187

9 Contract . . . 189
9.1 Introduction . . . 189
9.2 Adequacy of Basic Principles . . . 190
9.2.1 Freedom of Contract . . . 190
9.2.2 Freedom of Form . . . 191
9.2.3 Pacta Sunt Servanda . . . 193
9.2.4 Subjective Circumstances on the Part of AI . . . 194
9.3 Interpretation of Contracts Involving AI . . . 201
9.3.1 The Conclusion of the Contract by AI as a Circumstance Affecting the Interpretation of the Contract . . . 205
9.3.2 The Language of the Contract. In Dubio Contra AI . . . 206
9.3.3 AI as a Dominant Player in the Negotiations . . . 209
9.4 Due Diligence in Contracts Involving AI . . . 210
9.5 Performance of Contracts by AI . . . 211

9.6 Information Obligation . . . 216
References . . . 219
Books and Articles . . . 219
Documents . . . 220

10 Abuse of Right . . . 221
10.1 Introduction . . . 221
10.2 Abuse of Rights in the Context of the Principle of Respect for Human Autonomy . . . 224
10.3 The Abuse of Rights in the Context of Prevention of Harm . . . 226
10.4 Intellectual Advantage as an Abuse of Right? . . . 227
10.5 Tacit Collusion . . . 234
10.6 Conclusions . . . 237
References . . . 238
Books and Articles . . . 238
Documents . . . 238

11 Liability of AI . . . 239
11.1 Introduction . . . 239
11.2 Basic Concepts . . . 257
11.3 Legally-Relevant Damage Caused by AI . . . 264
11.4 Causation . . . 265
11.5 Negligence: Standard of Conduct (Reasonable Care, Due Diligence and so on)—The Novelty for AI or Also for Humans? . . . 269
11.6 Culpability of AI . . . 279
11.6.1 Legal Culpability for AI: Why Is It Needed? . . . 279
11.6.2 Legal Culpability: The Concept Representing Physical Fact or Social Fact? Is It Possible for AI to be Culpable? . . . 282
11.6.3 Legal Culpability: The Unified Concept, Radial Concept, or Many Concepts? Is It Possible at All to Cognize the Culpability of AI? . . . 285
11.6.4 Legal Culpability: An Autonomous or Relational Concept? How to Assess the Legal Culpability of an AI? . . . 287
References . . . 291
Books and Articles . . . 291
Documents . . . 293

12 Conclusions . . . 295

Chapter 1
Introduction: Is a New Conceptual Network Necessary to Adapt the Civil (Private) Law to the Development of AI and Robotics Development?

The growth of Artificial Intelligence (in the remainder of this book called "AI") and robotics in recent years has highlighted the pressing need to create a suitable legal framework. The debate on the subject is presently of quite general and preliminary character, despite many European acts and proposals for acts, and a plenitude of scientific books, reports and articles: its most important fields are being slowly defined, with the most pressing goal being the definition of the ethical foundations underpinning the further expansion of AI. In these preparatory works, there is a clear need to develop appropriate new civil law arrangements. Of all the branches of private law it is this one that has the greatest need for the settlement of new rules. Autonomous vehicles, medical robots, or expertise software demand essential questions on aspects of civil liability, such as culpability; in addition the growth in popularity of automated, intelligent software systems for concluding contracts requires a new approach to be taken to many fundamental and rooted contract law institutions, *inter alia* consciousness, intent, error, deception, interpretation of contract and good faith. Ruling on these specific matters demands the crystallisation of certain key points, which shall become the foundation for constructing a new AI/robot civil law. However, the current discussion on the civil law and AI is sketchy, superficial and lacks any reasonable order. A holistic coherent view on the issue of the civil law bases for the participation of AI in legal transactions is still lacking.

The first trial of this more comprehensive and wider-ranging debate on the civil law was initiated on the 16th of February 2017, by the legislative branch of the European Union, the European Parliament who called on the European Commission to elaborate new solutions based on civil (private) law that could respond to the rapid present-day development of robotics and AI.[1] For a long time, Resolution 2017 has

[1] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL), 2017 (called in the remainder of this book

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_1

represented the most significant declaration by the European community concerning future changes in civil law in this field.

Hence, the importance of Resolution 2017 for creating changes in civil law cannot be overestimated. It delineates a general framework within which experts should move, while also developing future aspects of civil law regulating the phenomena of AI and robots. Each element of the Resolution merits attention, because each could potentially have long-term implications regarding future legislation; furthermore the directions charted today will be difficult to change in future. In addition, Resolution 2017 also requested the Commission, on the basis of Article 225 of the Treaty on the Functioning of the European Union (TFEU), to submit, on the basis of Article 114 TFEU, a proposal for a directive on civil law rules on robotics that follows the specific recommendations set out in the Annex to the Resolution. Among other things, it insisted that the technological revolution should be shaped so that it serves humanity (the principle of beneficence), that the guiding ethical framework should be based on the principles and values enshrined in Article 2 TFEU and in the Charter of Fundamental Rights such as human dignity, equality, justice and equity, non-discrimination, informed consent, private and family life and data protection; it should also be based on other underlying principles and values of EU law, such as non-stigmatisation, transparency, autonomy, individual responsibility and social responsibility and on existing ethical practices and codes. Resolution 2017 required the protection of the safety, health and security of human beings, as well as of their freedom, privacy and integrity. It also recommended some technical means for realising these general principles, such as robot registers, compensation funds, individual registration numbers for robots and compulsory insurance. It should be noticed that in fact the request of European Parliament has been satisfied by the Commission only in part until now because the two key documents that were published by the Commission on the 28.09.22, did not answer all the European Parliament's requirements.[2]

The scope of application of Resolution 2017 has been purposely limited to intelligent robots, i.e., systems which can be characterised by the following attributes:

- the acquisition of autonomy through sensors and/or by exchanging data with its environment (inter-connectivity), and the trading and analysis of those data;
- self-learning from experience and by interaction (optional criterion);

"Resolution 2017"). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&rid=9, last access on the 4th of August 2022.

[2] Proposal for a directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM (2022) 496 final, embracing the fault liability regime, called in the remainder "Proposal ALD 2022". https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496. Proposal for a directive of the European Parliament and of the Council on liability for defective products, COM (2022) 495 final, embracing the strict liability regime, called in the remainder "Proposal DLDP 2022", https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A495%3AFIN&qid=1664465004344, last access on the 23rd of October 2022.

- at least minor physical support;
- the potential to adapt its behaviour and actions to the environment;
- absence of life in the biological sense.[3]

It should be noted that the distinction between AI and robots is not set in stone, and is unlikely to be maintained in the longer term; however, this does not present an obstacle for our analysis. Besides, in the course of our research, we arrived at the conclusion that the subject of future legislation should be Artificial Intelligence and not a robot.[4] It appeared that this direction of thinking accords with the legislative steps made by the European Union concerning AI. For instance, the European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence,[5] which is related to the part of civil law which is liability, used in the title the phrase "Artificial Intelligence". In addition, the Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts[6] explicitly uses the phrase "Artificial Intelligence" and provides a new definition in the Article 3 (1):

> 'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;

This means that the main subject of legislation now is not a robot but AI. Although we recognize that differences exist between robot and AI, this monograph uses the names "AI" and "robot" interchangeably, and sometimes, more loosely "machine", for the sake of style, where the difference is not important.

In his famous and insightful book *The Structure of Scientific Revolutions*, T. Kuhn writes about the invisibility of scientific revolutions, which are usually viewed not as revolutions, but as additions to scientific knowledge. The main reason of this regularity is that

> Both scientists and laymen take much of their image of creative scientific activity from an authoritative source that systematically disguises [...] the existence and significance of

[3] The Resolution 2017, Point 1.

[4] Księżak and Wojtczak (2020).

[5] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)), P9_TA(2020)0276, called in the remainder "Resolution 2020". https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html, last access on the 4th of August 2022.

[6] Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, 21.04.2021, COM (2021) 206 final called in the remainder "Proposal 2021"). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206, last access the 4th of August 2022.

> scientific revolutions. [...] both the layman's and practitioner's knowledge of science is based on textbooks and few other types of literature derived from them [...][7]

Meanwhile, as Kuhn reports, the textbooks show the endeavour of science as a long-standing historical tradition. We believe that the same is even more true in relation to law and legal science. Because stability is seen there as an especially important positive value, the authors of normative acts, commentaries, monographs and textbooks wish to continue the conceptual network grounded so far, or at most merely develop it. They forget, ignore or even do not know, that sometimes such continuation is not possible and not positive. When the reality external to the law, especially in a part regulated by the given legal discipline, is a matter of diametrical change, the legal discipline must change diametrically with it. Kuhn observes:

> Many puzzles of contemporary normal science did not exist until after the most recent scientific revolution. [...] Earlier generations pursued their own problems with their own instruments and their own canon of solutions.[8]

And so, he insists:

> Textbooks [...] have to be rewritten in whole or in part whenever the language, problem-structure, or standards of normal science change. In short, they have to be rewritten in the aftermath of each scientific revolution [...][9]

Applying this phrase analogically to law and legal sciences, we must remember that its author did not assume a normative perspective. He merely described some regularities in the historical processes taking place in Science. And one should treat another of his important thesis in the same way, that rewriting the textbooks does not mean making the change clear and explicit.:

> [...] once rewritten, they inevitably disguise not only the role but the very existence of the revolutions that produced them. [...] Since new paradigms are born from old ones, they ordinarily incorporate much of the vocabulary and apparatus [...] that the traditional paradigm had previously employed. But they seldom employ these borrowed elements in the traditional way. Within the new paradigm, old terms, concepts, and experiments fall into new relationships one with another. The inevitable result is what we must call, though the term is not quite right, a misunderstanding between the two competing schools. [...] Before those [postrevolutionary] texts are written, while debate goes on [...] the opponents of a new paradigm can legitimately claim that even in the area of crisis it is little superior to its traditional rival. [...].[10]

The aim of the analogy presented above is to show that jurists, both practitioners and scientists, struggle with the tendency to react to the development of AI and robotics as they usually did or as they think they did—it is step by step and bit by bit. They try to apply to the new situation old schemes, axioms and mental images. Meanwhile, in response to the revolutionary and massive changes occurring in the world of

[7] Kuhn (1996), pp. 136–137.

[8] Kuhn (1996), p. 140.

[9] Kuhn (1996), p. 137.

[10] Kuhn (1996), pp. 137, 149–156.

technology, and the resulting changes in everyday life, the law and legal sciences must find new concepts, ideas and instruments useful to resolve new problems. These concepts, ideas and instruments should not be hidden under the old ones to give the impression of continuity of legal tradition: the cost of such continuity may be too high. The jurists must rewrite their textbooks in a way that does not intentionally disguise what has happened. Only by not burying their heads in the sand will they avoid misunderstandings and taking an overdue, improper or inadequate steps.

The title of this book is a meaningful one and has arisen as a result of the differentiation between terms and concepts which is strongly internalized by the authors. Terms are names or labels we give to physical or abstract objects. Their function, besides the those of economics of communication and cognition, is to make the ostensive definition of the physical object or a concept possible.[11] In turn, a concept is the set of beliefs and postulates about a characteristic of a physical or abstract object. The title of our book *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, means that we do not study terms but concepts. We are convinced that the problems arising from many of the debates taking place today around the law and AI are caused by a lack of adherence to the differentiation between terms and concepts. People who use the same terms for different concepts or contrariwise, or different terms for the same concept, are comparing apples and oranges, as vividly encapsulated by Pagallo (2018a, b).

It is not a new problem: when changing life circumstances demand changes in reasoning and language in the aim to describe the world adequately, the question arises of whether to give to the new phenomena entirely new names identifying their new concepts, or to use old names but modifying and broadening simultaneously their concept.[12] The former method was chosen in Poland in 2009 when the English term "leasing" was incorporated in the code next to the Polish term "najem", a traditional rental agreement, in the kinds of contracts regulated by the Polish Civil Code. The problem with this method is that the number of concepts is multiplied, and the law may become incomprehensible for people. Regarding the latter method, including representations of new phenomena into the image of the world usually uses metaphorical projection as a tool justifying the change of old concepts; as a result the category described by the old concept becomes a radial category.[13] For example in the Polish Criminal Code of 1969, the offence of handling regarded only tangible things; however, in the Criminal Code of 1997, the concept of handling was broadened to include the computer software by supplementing the traditional concept of handling with that of handling of computer software. The problem with the latter method is more significant for the debates *de lege ferenda* because it increases the number of misunderstandings caused by the equivocation fallacy. The best instances of such misunderstandings are the discussions on the legal subjectivity

[11] Cf. the remarks of Wittgenstein on ostensive definition—Wittgenstein (2009), § 30 and 258.

[12] On broadening the legal concepts cf. Wojtczak (2013).

[13] On metaphor as a tool of broadening the legal concepts cf. Wojtczak (2017).

or legal personality for AI, on the capability of concluding contracts, and on the attributability of responsibility.

There are several things that should be settled before the topic will be elaborated.

Firstly, this monograph is not a review, therefore, its aim is not to summarize the current *status quaestionis* on the topic. We do not want to survey and summarize previously published literature on AI and law, but to present our own concepts and ideas, of course based on up to day *aquis academique*. This assumption is almost necessary because today the scientific and the popular culture is burgeoning with a continual deluge of books and articles of varying value, and it is not possible even to register them all. Even so, we are convinced that this effort by many authors, and our own, is not a useless one, as a reference or even in the mind. We are all bearing witness to an intense global debate on the shape of a future world where people coexist increasingly closely with new technologies. And this book is intended as an element of this debate.

Secondly, if it is to live for more than a year, our book must be a little futuristic. Technology changes so fast that it cannot be predicted exactly what new technical solutions will operate on the market in two or more years. We can only suspect that they will be more complicated, more powerful, more autonomous and, what is of great importance, increasingly omnipresent in everybody's everyday life. Hence, we as a rule, do not accept the criticisms based on arguments of the type "it (technology, situation, idea, etc.) is not possible". Simultaneously, we limit our interest to the weak (narrow) AI, although it may be stronger than the one today. The monograph does not research on the consequences of superintelligence[14] for the civil law, because in our opinion, as in the Verge's concept of the singularity,[15] it is not possible to predict what would happen then.[16]

Thirdly, the subject of our research is the civil (private) law, mainly that of continental Europe, based on the Roman law tradition. Despite this, because we are interested in the general core of normative institutions and constructions, and not in the differences in these institutions and constructions on the lower level, our concepts and ideas are not placed in any concrete legal system. Hence, we also make use of acts such as DCFR,[17] PECL[18] and Unidroit.[19] We want to examine the influence of the growth of AI on the fundamental civil law concepts common to

[14] On the general consequences of superintelligence Bostrom (1998).

[15] Verge (1993).

[16] As Mahler (2022), p. 521 rightly points out "[D]ystopian superintelligence scenarios are highly controversial and uncertain so regulating existing narrow AI should be a priority."

[17] Study Group on a European Civil Code & Research Group on EC Private Law (Acquis Group), Principles, Definitions and Model Rules of European Private Law. Draft Common Frame of Reference. Outline Edition, Sellier European Law Publishers: Munich 2009.

[18] The Principles on European Contract Law, https://www.trans-lex.org/400200/_/pecl/, last access the 4th of August 2022.

[19] Principles of International Commercial Contracts, International Institute for the Unification of Private Law, Rome, February 2004, https://www.unidroit.org/english/documents/2004/study50/s-50-98-e.pdf, last access the 4th of August 2022.

all legal systems belonging to the Western tradition, especially those of continental Europe.

Fourthly, we do not refer to particular problems of the constitutional, criminal or administrative law, insofar they are not related directly to the private law. Nevertheless, of course, we make some strong assumptions about them because legal systems are internally strongly interrelated and integrated. One of such assumption of a constitutional character, is that we research on the AI civil law based on a democratic liberal form of government. We are rather not interested in legal solutions introduced in such political regimes like that of China, although they may be economically very effective. We also assume that it will be necessary to regulate some issues about AI on the constitutional level.[20] Furthermore, we assume that the main part of AI regulation will be situated within the scope of the administrative law, and that the scope of this law will often be supranational: regional, European or global. The main topics there will be classification of AI and robots, certification and registration systems. The initial rules, and many postulates, regarding classification, certification and registration are already made. The most obvious are classifications of robots according to their function, for example distinguishing such devices like unmanned aircrafts,[21] medical devices[22] or autonomous weapon systems.[23] Another set of rules and postulates are these regarding registration. For example, in Resolution 2017 point 1, the European Parliament in states that:

> [...] a comprehensive Union system of registration of advanced robots should be introduced within the Union's internal market where relevant and necessary for specific categories of robots, and calls the Commission to establish criteria for the classification of robots that would need to be registered; in this context calls on the Commission to investigate whether it would be desirable for the registration system and the register to be managed by a designated EU Agency for Robotics and Artificial Intelligence.

Then in point 59.e, the European Parliament calls on the Commission to explore the implications of such legal solution as:

> ensuring that the link between a robot and its fund would be made visible by an individual registration number appearing in a specific Union register, which would allow anyone interacting with the robot to be informed about the nature of the fund, the limits of its

[20] Księżak (2021).

[21] Cf. Commission Implementing Regulation (EU) 2019/947 of 24 May 2019 on the rules and procedures for the operation of unmanned aircrafts L 152/45. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R0947, last access on the 4th of August 2022; Commission Delegated Regulation (EU) 2019/945 of 12 March 2019 on unmanned aircrafts systems and on third-country operators od unmanned aircraft systems. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019R0945, last access on the 4th of August 2022.

[22] Cf. Directive 93/42/EEC concerning medical devices modified by Directive 2000/70/EC (MDD)). https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1993L0042:20071011:en:PDF, last access on the 4th of August 2022.

[23] Cf. European Parliament resolution of 12 September 2018 on autonomous weapon systems (2018/2752(RSP)). https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341_EN.html, last access on the 4th of August 2022.

> liability in case of damage to property, the names and the functions of the contributors and all other relevant details[...].

A further example is European Commission, White Paper: On Artificial Intelligence—A European approach to excellence and trust, Brussels, 19.2.2020,[24] which in point 5.D.b declares:

> [...] the regulatory framework could prescribe that the following should be kept:
> - accurate records regarding the data set used to train and test the AI systems, including a description of the main characteristics and how the data set was selected;
> - in certain justified cases, the data sets themselves;
> - documentation on the programming and training methodologies, processes and techniques used to build, test and validate the AI systems, including where relevant in respect of safety and avoiding bias that could lead to prohibited discrimination.
>
> The records, documentation and, where relevant, data sets would be needed to be retained during limited, reasonable time period to ensure that they are made available upon request [...]

The postulates are made concrete by the proposal of the system of registration for high-risk AI given in Proposal 2021.

We agree that a special system of registration and certification should be introduced within the scope of administration law, however, that even if the general register, be it global, European or state level, is dedicated only for the riskiest or most influential AI systems, the less significant, i.e., regional, local or industry registers should be dedicated for other AI systems, so that no AI would stay outside the registration/certification system; this would prevent any AI falling outside the registration/certification procedure. AI systems which are not registered or certificated should be acknowledged as illegal and should be eliminated out of the market.

The next, sixth, general assumption made during the writing of this book is that the real field we need to address is not that of robots, but generally AI. This is in accordance with the direction of debate in the European Union, which first moved towards the postulate of Civil Law Rules on Robotics and the regulation for robots but was then redirected to AI and AI systems.[25] We are aware, of course, of the problems with AI definition,[26] but we believe that they should be solved *ad hoc* as the specific regulations are created.

To conclude our presentation of our the general assumption, we must once more insist that our purpose is to discuss the general conceptual network of the civil law in the context of AI and emerging technologies. Therefore, we do not address the more detailed problems connected with specific economical market domains, such as antitrust law or public contracts, or on the problems of detailed EU legislation, which mostly serve as a form of changeable "overlay" on more traditional civil law concepts.

These limitations of the scope of our book are unfortunately necessary because writing about AI is a very thankless task. Bertolini (2020), p. 15 captured this

[24] Called in the remainder "White Paper 2020".

[25] Księżak and Wojtczak (2020).

[26] For the description of the currently used definitions of AI cf. Bertolini (2020), pp. 15–21.

problem in a very pertinent phrase: "AI is a moving target: what is deemed an AI application is no longer considered as such when technology advances".

What points should be considered just to rewrite the textbooks? These points determine the structure of this book. It consists of 12 chapters, making historical accounts only when it is necessary for the commented topic.

The second chapter presents the need to reflect on the concept of legal personhood and legal personality as the key concept of the private law constituting other concepts. We will try to answer the following questions: Is it possible for AI to have legal personality and is it useful? On what conditions can it exist? Is AI a philosophical zombie[27]? If so, what are the consequences for private law and the related concepts?

The third chapter focuses on AI's will and discernment. Do they exist? If so, are they the same as the human ones? Are they necessary for the participation of AI in legal transactions?

The fourth chapter considers the capability for juridical acts of AI. It analyses the following problems: the relationship between the capabilities of natural persons and possible capabilities of AI; and the similarities and differences between the capabilities of legal persons and capabilities of AI; is AI more like a legal person or like a juvenile or maybe we need a different perspective. This part of the monograph also demonstrates why traditional legal concepts of judgment and will are inadequate within the world of AI. The authors believe that these traditional concepts are wrong, even for natural persons.

In the following step, and it is placed in the fifth chapter, the concept of consent is analysed. In what legal sense it is possible to speak about the consent of AI? Are there any features that AI should have, to receive effectively the consent of another?

The sixth chapter examines personal interests: are the personal interests of AI possible and useful, and on what conditions is this so? It also examines how this question may be related to the concept of personality of AI, and the possible kinds of personal interests of AI.

The seventh chapter discusses the theses concerning intellectual property, with the following specific themes: AI as a piece of work, AI as an actual and legal author of a piece of work, the holder of immaterial (moral) rights to the piece of work made by AI and the holder of property rights to it.

The eight chapter indicates which changes in the concept of property are necessary when the AI may occupy the following positions in legal relationships: AI as a property AI as a possessor of property, AI as an owner of property.

The ninth chapter attempts to explain some specific problems arising when contracts are concluded by AI and the subjects of law or when both parties are AI.

The tenth chapter tries to put the problem of the abuse of right: what new kinds or ways of abusing rights are possible in the world of AI and do these require changes of the traditional concept of abuse of rights.

[27] Cf. Chalmers (1996), p. 94.

The nine chapters examine more rudimentary problems and provide a background to analyse the issue of liability. The eleventh chapter starts with a short summary of existing rules, proposals, and basic concepts necessary to examine the problems of AI's liability. This is followed by an examination of the legally-relevant damage caused by AI, and then is a discussion of causality, focusing on the traditional legal concept of causality in private law as an element of deciding on AI's liability. Finally, the chapter reflects on negligence and culpability (fault) when AI is involved are reflected on. It makes particular mention of determining the standard of reasonable care imposed on AI and on human beings living in the world where AI is present.

The whole of the monograph is ended with the conclusions.

The structure of the book itself is aimed to advance the discussion of the status of AI in legal transactions. Even if our proposals of concrete legal regulations eventually turn out to be incorrect or unacceptable, we hope that they will give rise to new ideas or solutions among critics. This form of debate seems to be necessary because, in fact, the existing moot points are not known or given in a conclusive way. The future, even a relatively near one, is difficult to foresee. However, a lack of perfect clarity cannot force us to stop and wait for some conclusive knowledge to reveal itself. By delaying, as it is said by some pessimists, we risk losing the chance to control the growing battle for the power over the world.

References

Books and Articles

Bostrom N (1998) How long before superintelligence? Int J Future Stud 2. https://nickbostrom.com/superintelligence.html, last access on the 4th of August 2022

Chalmers DJ (1996) The conscious mind: in search of a theory of conscious experience. Oxford University Press, New York

Księżak P (2020) Zawieranie umów przez sztuczną inteligencję (AI). In: Dumkiewicz M, Kopaczyńska-Pieczeniak K, Szczotka J (eds) Sto lat polskiego prawa handlowego. Księga jubileuszowa dedykowana Profesorowi Andrzejowi Kidybie, vol. II. Wolters Kluwer, Warszawa

Księżak P (2021) My, Naród? Konstytucjonalizacja sztucznej inteligencji, czyli o potrzebie przemodelowania założeń ustrojowych. Przegląd Sejmowy 4(165):65. https://doi.org/10.31268/PS.2021.46

Księżak P, Wojtczak S (2020) AI versus robot: in search of a domain for the new European civil law, Law, Innovation and Technology. Taylor & Francis Online https://doi.org/10.1080/17579961.2020.1815404

Kuhn T (1996) The structure of scientific revolutions. The University of Chicago Press, Chicago

Mahler T (2022) Regulating artificial general intelligence (AGI). In: Custers B, Fosch-Villaronga E (eds) Law and artificial intelligence. Regulating AI and applying AI in legal practice. Asser Press-Springer, Cham, Switzerland

Pagallo U (2018a) Vital, Sophia, and Co. The quest for the legal personhood of robots. Information (Switzerland) 9(9):230. https://doi.org/10.3390/info9090230

Pagallo U (2018b) Apples, oranges, robots: four misunderstandings in today's debate on the legal status of AI systems. Philos Transact Royal Soc A 376:20180168. https://doi.org/10.1098/rsta.2018.0168
Verge E (1993) The coming technological singularity: how to survive in the Post-Human Era. https://edoras.sdsu.edu/~vinge/misc/singularity.html, last access on the 4th of August 2022
Wittgenstein L (2009) Philosophical investigations (trans: Anscombe GEM, Hacker PMS, Schulte J). Blackwell, Oxford. Revised 4th edition by Hacker P M S and Schulte J
Wojtczak S (2013) The broadening of legal notions as a tool in the neutralization of values in law. In: Pałecki K (ed) (2013) Neutralization of values in law. Warszawa, Wolters Kluwer
Wojtczak S (2017) The metaphorical engine of legal reasoning and legal interpretation. C.H. Beck, Warszawa

Documents

Bertolini A (2020) Artificial Intelligence and Civil Liability. Study. Requested by the European Parliament's Committee on Legal Affairs. July 2020. Brussels. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf, last access on the 4th of August 2022
Commission Delegated Regulation (EU) 2019/945 of 12 March 2019 on unmanned aircrafts systems and on third-country operators od unmanned aircraft systems, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019R0945, last access on the 4th of August 2022
Commission Implementing Regulation (EU) 2019/947 of 24 May 2019 on the rules and procedures for the operation of unmanned aircrafts L 152/45, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R0947, last access on the 4th of August 2022
Directive 93/42/EEC concerning medical devices modified by Directive 2000/70/EC (MDD). https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1993L0042:20071011:en:PDF, last access on the 4th of August 2022
European Commission. White Paper On Artificial Intelligence. A European approach to excellence and trust COM (2020) 65 final. 19.2.2020. Brussels. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf, last access on the 4th of August 2022
European Parliament resolution of 12 September 2018 on autonomous weapon systems (2018/2752 (RSP)). https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341_EN.html, last access on the 4th of August 2022
European Parliament resolution of 16 February 217 with recommendations on the Commission on Civil Law Rules on Robotics (2015/2103(INL)), P8_TA (2017)0051. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&rid=9, last access on the 4th of August 2022
European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)), P9_TA(2020)0276. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html, last access on the 4th of August 2022
Principles of International Commercial Contracts, International Institute for the Unification of Private Law, Rome, February 2004, https://www.unidroit.org/english/documents/2004/study50/s-50-98-e.pdf, last access on the 4th of August 2022
Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, 21.04.2021, COM (2021) 206 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206, last access on the 4th of August 2022

Study Group on a European Civil Code & Research Group on EC Private Law (Acquis Group), Principles, Definitions and Model Rules of European Private Law Draft Common Frame of Reference. Outline Edition, Sellier European Law Publishers: Munich 2009 (DCFR). https://www.law.kuleuven.be/personal/mstorme/european-private-law_en.pdf, last access on the 4th of August 2022

Chapter 2
Artificial Intelligence and Legal Subjectivity

2.1 Introduction

In Resolution 2017 the European Parliament called on the Commission to explore, analyse and consider the implications of legal solutions, particularly those creating a specific legal status for robots in the long run. These solutions would at least establish the most sophisticated autonomous robots as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third party independently (point 59f of Resolution 2017). The position of the European Parliament generated different opinions. For example, the European Economic and Social Committee in the opinion accepted 3 months later and entitled *The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society* in § 3.33 announced that

> The EESC is opposed to any form of legal status for robots or AI (systems), as this entails an unacceptable risk of moral hazard. Liability law is based on a preventive, behaviour-correcting function, which may disappear as soon as the maker no longer bears the liability risk since this is transferred to the robot (or the AI system). There is also a risk of inappropriate use and abuse of this kind of legal status. The comparison with the limited liability of companies is misplaced, because in that case a natural person is always ultimately responsible.

Similar objections were expressed in the *Open Letter to the European Commission: Artificial Intelligence and Robotics* (http://www.robotics-openletter.eu, last access on the 4th of August 2022) signed in April 2018 by 285 political leaders, AI/robotics researchers, industry leaders, physical and mental health specialists, law and ethics experts.

The main part of this chapter is based on Wojtczak (2022).

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_2

At last, influenced by these disputes EP changed its opinion and on the 20th October 2020 in Resolution 2020 (Annex. B.(6)) stated:

> Any required changes in the existing legal framework should start with the clarification that AI-systems have neither legal personality nor human conscience, and that their sole task is to serve humanity. Many AI-systems are also not so different from other technologies, which are sometimes based on even more complex software. Ultimately, the vast majority of AI-systems are used for handling trivial tasks without or with minimum risks for the society.

And in the same date in the European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies (2020/2015(INI)) point 13 the European Parliament

> Notes that the autonomisation of the creative process of generating content of an artistic nature can raise issues relating to the ownership of IPRs [intellectual property rights] covering that content; considers, in this connection, that it would not be appropriate to seek to impart legal personality to AI technologies and points out the negative impact of such a possibility on incentives for human creators.

These controversies are one of the factors which influence the content of this chapter, which reflects on the problem of legal subjectivity or legal personhood for AI. Another factor is that solving this problem is conceptually necessary for all the issues researched in this book. Nothing can be said of contract and tort law for AI, or property law and intellectual property law, without making some stable assumptions about the legal subjectivity or legal personhood of AI. It is first necessary to make two important remarks. First, we are not interested in the normative aspect of the discussion, so we do not discuss who or what deserve full or partial legal subjectivity or personality. Only the descriptive image of the concept is important here: who or what is used to be a legal subject or legal person, and what can be predicated on this basis as to the future legal subjectivity or legal personhood of AI. Second, we are aware that for various reasons, it is often insisted that legal personhood is differentiated from legal subjectivity,[1] but simultaneously we are convinced that they are interconnected strongly, especially on the grounds of civil law, while both of them are used in a very random manner. There are, for example, legal systems, for example the Polish one, where the term "legal person" is used in legislative acts in a very narrow meaning: in the Polish Civil Code, the term applies for certain kinds of companies or institutions but not for human beings, which are called natural persons. But such a use of this term and the sense ascribed to it can be treated as belonging to the language of the concrete legislative acts, not covering the entire scope of the concept used in by the jurists. Usually, the theory of civil law and jurisprudence in general indicate that "legal personhood" has a broader meaning, and we agree with this opinion. When the concept of legal subjectivity presented below is accepted, this notion looks like indistinguishable from that of legal personhood. Taking into consideration the fact that even non-living objects like companies can be called "legal persons", it is difficult to maintain that this name accrues only to the philosophical or moral persons, i.e. to human beings alone. Therefore, treating

[1] Such differentiation is described by Pietrzykowski (2018), pp. 7–23.

interchangeably legal personhood and legal subjectivity does not entail accepting the philosophical or ethical position that moral subjectivity is that same as moral personhood. Since it should be remembered that in law, as observed Naffine (2009), pp. 1, 3:

> The 'person' is the formal subject of rights and duties: a legal idea or construct, not to be mistaken for a real natural being. The legal use of the term 'person' therefore should not be taken to entail any larger biological, philosophical or even religious claim or implications [. . .]. Law has no one type person in mind.

It is rightly observed by Bryson et al. (2017), p. 280 that *[l]egal personhood is not an all-or-nothing proposition*. Legal subjectivity is a complex attribute which may be recognized in certain entities or assigned to others. This attribute is, in our opinion, gradable, discrete, discontinuous, multifaceted and fluid. It means that it can consist of more or fewer elements of different types (e.g., responsibilities, rights, competences, and so on), which can be added or taken away by a lawmaker in most cases; the exception being human rights, which, according to the prevalent opinion, cannot be taken away. Among others, such character of this attribute can be seen in the contemporary Polish civil law doctrine, which distinguishes the following concepts determining subjectivity:

- natural persons, i.e. human beings (Article 8 § 1 of Polish Civil Code),
- legal (juristic) persons, i.e. the State Treasury and organizational entities in which specific provisions vest legal personality (Article 33 of Polish Civil Code)
- defective juristic persons (so-called), i.e. organizational entities that are not legal persons, in which a statute vests legal capacity (Article 331 § 1 of Polish Civil Code) and
- other entities, i.e. those not classified as any type of persons but endowed with some claim-rights, responsibilities, and/or competences, e.g. animals which, according to Article 1 § 1 of Polish Animals' Protection Act (1997), are regarded as living entities but not things or persons, and which are able to feel pain, and are entitled to respect, protection and care.

Therefore, while we do not agree with all its details, we accept the general spirit of the Bundle Theory of Legal Personhood proposed by Kurki (2019), which is based on two key tenets:

1. Legal personhood of X is a cluster property and consists of incidents which are separate but interconnected.
2. These incidents involve primarily the endowment of X with particular types of claim-rights, responsibilities, and/or competences.[2]

We are convinced that the concept of legal subjectivity itself is open-ended, defeasible and ascriptive in a Hartian sense.[3] It can be paraphrased that:

[2] Kurki (2019) uses the typology of Hohfeld (1920).

[3] Hart (1948–1949).

> Our concept of an action [here – a legal subjectivity], like our concept of property is a social concept and logically dependent on accepted rules of conduct. It is fundamentally not descriptive, but ascriptive in character; and it is a defeasible concept to be defined through exceptions and not by a set of necessary and sufficient conditions whether physical or psychological.

Hence, when a lawmaker or a judge claims: "He/it is a legal subject" they do not describe anything. Instead, they use an ascriptive. This is a very clear approach, well-fitting to the concept of legal subjectivity:

> Like directives, ascriptives are attempts by the speaker to get the listener to do something. [. . .] Like commissives, ascriptives are those illocutionary acts which point is to commit the speaker to some future cause of action. Saying "Guilty", the judge determines not only the new legal position of the listener but indicates that he or she must be responsible for an offence or misdeed. [. . .] As well as declarations, ascriptives presuppose the existence of extralinguistic conventions necessary for success of this type of the speech act that postulate the special social statuses of speaker and listener.[4]

It is possible to think of subjectivity, especially legal subjectivity, at least in three ways: (1) philosophically, (2) from the perspective of law in general, (3) from the perspective of a law that is valid in a certain place and a certain time. However, it is important to note that these perspectives do not simply refer to the same object viewed in terms of the most general to most specific: instead, they represent three different kinds of thinking and concern different objects. These kinds of thinking are often confused, so it is important to avoid falling into this trap.[5] Of course, as all three ways of understanding subjectivity influence culture, they also influence each other. The first one can be regarded as religious thinking, in the sense of Finnis (2011),[6] and is often connected with moral subjectivity. The second relates to the subjectivity present in law, but not in *the law*;[7] its overlap with the demands of the law of a given country remains a matter of controversy. However, such controversy generally remains unnoticed; it only becomes significant, and of practical value, in times of crisis, especially political or humanitarian ones.[8] The third is purely juristic: it relates only to the concept used in the acts and doctrine of concrete

[4] Ogleznev (2016).

[5] Kurki (2019) is correct to notice that "theories of corporations are thus often combinations of social ontology, normative political philosophy, and analytic jurisprudence".

[6] Finnis (2011) uses the word "religion" to name the basic value of reflections on the origins of cosmic order, human freedom and reason, universal order of things, etc., by a human being, regardless of the answer to the given question.

[7] Finnis (1987) describes this difference in the following way: "Positivists and natural law theories in jurisprudence are not, and do not even look like, theories about *the law* of any particular community (in the sense of offering to identify propositions of law which are true for that legal system), or about the criteria for identifying the law which are used by the lawyers and judges of any particular community. They look like theories about what *law* – a(ny) legal system – 'necessarily is' (at least in its paradigmatic instantiations, its central cases)".

[8] We mean the reflections which refer to some essential or *a priori* concepts of law. For example, Radbruch (1945).

legal systems, as well as in the Resolution 2017, Resolution 2020, the Proposal 2021 and other documents of European law.

For further discussion it is necessary to remember that the concepts of legal subjectivity and legal personhood, analysed from the perspective of a concrete legal system, are dependent on two key institutions. The first institution is a legal capability, understood as being the subject of rights and duties; however, this is independent of the capability to act on one's own behalf. The second institution is the capability to perform legal acts, understood as the capability to conclude legal acts with one's own actions. In the traditional view, the entity may be endowed with legal capability, being not able to conclude legal acts on its own behalf.

2.2 Alleged Hierarchy: (AI?), Human Beings, (AI?), Juristic Persons, (AI?), Animals

In the literature, two key analogies are used when discussing the possibility of acknowledging legal subjectivity or legal personhood for AI systems: one between AI and animals, and another between AI and juristic persons or collective subjects.[9] Many researchers agree that legal subjectivity in the form acknowledged to a human being is unique and cannot be acknowledged to AI, especially because, for now at least, AI does not demonstrate any evidence of being conscious and sentient.[10] Many researchers want to make legal subjectivity dependent on the moral status of entity. In 2017 UNESCO World Commission on the Ethics of Scientific Knowledge and Technology published *Report of COMEST on robotics ethics*. In sections 201 and 202, it is stated that:

> From a deontological point of view, to have moral status implies being a person, and being a person implies having rationality or the capacity for rational and moral deliberation. In so far as they are able to solve many demanding cognitive tasks on their own, robots may be said to have some form of rationality. However, it is highly counterintuitive to call them 'persons' as long as they do not possess some additional qualities typically associated with human persons, such as freedom of will, intentionality, self-consciousness, moral agency or a sense of personal identity. [...] From a utilitarian perspective, moral status does not depend on rationality, but on sentience or the capacity to experience pleasure and pain (broadly construed) and the accompanying emotions. According to this view, humans and many non-human animals have moral status, but robots do not, because they are not sentient and

[9] Solaiman (2017), Chen and Burgess (2019), and Kurki and Pietrzykowski (2017).

[10] However, when employing philosophical reflection, some writers insist on the necessity of rejecting naïve humanism, as well as the belief that human claims to subjective treatment are exclusive, and that entities which fulfill the biological criteria of humanity enjoy a privileged moral status. Hence Pietrzykowski (2018) based on the collective criteria of psychological abilities, genetic, morphological and anatomical attributes, proposes that personal subjects (including human and other persons) should be differentiated from non-personal subjects (including human non-personal subjects and extra-human non-personal subjects).

lack emotions. According to some authors (e.g. Torrance 2013), genuine sentience can be ascribed only to organic beings, not to robots.

In contrast, analogies with animals appear more suitable, as the abilities of AI are limited in relation to humans. On the other hand, AI can be regarded as analogous to collective entities in the sense that it is an artificial creation, a non-biological one lacking in sensations and consciousness. Besides, according to the traditional Western view, animals and juridical persons are, next to humans, the only true candidates for broader- or narrowly-determined legal subjectivity. Many Western jurists would be surprised to learn that in some countries or cultures, rivers have also been acknowledged as subjects of law, such as the Ganges Jamuna in India and the Whanganui in New Zealand.[11]

However, using an analogy with animals or juristic persons to justify awarding potential legal subjectivity to AI requires a certain superficial assumption. Firstly, this analogy assumes that there is a single hierarchy or sequence of entities, organized according to their degree of similarity to human beings,[12] and, secondly, that the place of an entity in this hierarchy or sequence (based on the degree of development) determines the scope of subjectivity attributed to it. It follows that animals take the lowest place in the hierarchy, because despite being endowed with sentience, they lack reason, which is traditionally regarded as an essential and uniquely human feature. In the same way, the next place could be taken by contemporary AI, which lacks sentience and its reason is not perfect. The next position up the hierarchy is taken by collective entities, because they lack sentience but have collective reason; such reason corresponds to, and may surpass, human reason because its substrate is human. Finally, at the top of the hierarchy are human beings; these are sentient and have reason which is, according to traditional views, the best, prototypic example of its kind.

Taking this way of thinking, it can be anticipated that if AI developed to such an extent that it could achieve complete reason, or a form superior to human reasoning, and if it gained some sentience, it would be elevated above collective entities and be ranked on par with human beings.[13] The advocates of this vision believe that AI cannot gain a different legal subjectivity to that enjoyed by animals or collective

[11] Kowalski (2017) in a serious, juridically-profiled Polish daily newspaper asked: "Contemporary lawmakers start to respect the archaic Maori point of view. What is happening when the modern way of thinking accepts an archaic mentality? [. . .] Watching the actions of India and New Zealand governments it is not possible to forget that these countries are situated very far from Poland not only in geographic sense, but also a mental one".

[12] Chen and Burgess (2019), pp. 79–80 say: "Human beings are in many ways, the default position in relation to legal personhood [. . .] Legal personhood is, however, also extended to other entities that are not humans [. . .] One of the more basic and common instances of this recognition relates to the corporate structure. [. . .] In recent decades, arguments have also been made to extent legal personhood to non-human animals. [. . .] Other arguments can be made for the extension of legal personhood to other entities that have been created by humans".

[13] It is difficult to imagine what the world would be like if creatures endowed with reason, conscious and sentience better than humans were to exist. What place would they take in the legal hierarchy? It is possible that, as in the Watts (2006) novel "Blindsight", where humans live alongside creatures

entities: it must be similar, derivative. So, if in a given legal system, collective entities may have some rights specific for humans but not attributed to animals (e.g., personal goods/rights), which were awarded on the basis of similarity; in such a system, an AI is acknowledged as being higher than animals may also be endowed with similar rights as humans. A good example of this kind of thinking is illustrated in copyright law: it has been argued that AI cannot be acknowledged as an author based on various cases related to animals, especially the famous case of the monkey's selfie (Naruto v. Slater, No. 16-15469 [9th Cir 2018]). This case will also be discussed briefly below.

In fact, this hierarchical and analogical conception is proposed by COMEST in the sections 203–205 of the report, although ostensibly it can look similar to the concept presented below. The main reason of the failure of the COMEST's proposal is that it is made from the moral and psychological perspective and not exclusively from the social or the legal one. COMEST says:

> A possible third way of assigning moral status to robots (a way that does not focus on any particular psychological or biological property) is to adopt a 'relational perspective', according to which robots would possess moral status in so far as they participate in unique, possibly existentially significant, relationships with human beings. [. . .] When it comes to robots, however, this 'relational' solution could face the problem of depending on the human psychological tendency to anthropomorphize or 'project' human properties onto inanimate objects and artefacts. [. . .] Yet, this bonding between humans and robots is not necessarily the result of antropomorphization: also without being human-like, technological artefacts like robots can become so meaningful and valuable that they deserve to be protected [. . .] The rapid development of highly intelligent autonomous robots, then, is likely to challenge our current classification of beings according to their moral status, in the same or maybe even more profound way as it happened with non-human animals through the animal rights movement.

The failure of this idea lies in taking the moral point of view which pushes the COMEST's conception into the sphere of intrinsic and hardly irrefutable values. This may in fact incur dangerous consequences, especially for the human situation in the world. This danger becomes clear when the next part of this section is read:

> It may even alter the way in which human moral status is currently perceived. Although still resembling futuristic speculations, questions like these should not be dismissed lightly, especially in view of the fact that the 'human-machine divide' is gradually disappearing [. . .] and the likelihood of future appearance of human-machine or animal-machine hybrids or cyborgs [. . .].

2.3 Sentience and Reason

In addition to the objection expressed above, there are several other reasons why the theories described above should be rejected as mistaken.

---

which are genetically reincarnated vampires, the two races would be locked in an ongoing trial of power or even a battle for supremacy and leadership.

To clarify, it would first be useful to remind ourselves of an argument based on a well-known mental experiment by Chalmers, assuming logical possibility of the existence of a zombie. What is a zombie? According to Chalmers (1996), p. 84 it is

> someone or something physically identical to me (or to any other conscious being) but lacking conscious experiences altogether. [...] This creature is molecule-for-molecule identical to me, and indeed identical in all the low-level properties postulated by a completed physics [...] He will certainly be identical to me functionally [...]. He will be psychologically identical to me [...]. He will even be "conscious" in the functional senses [...] – he will be awake, able to report the contents of his internal states, able to focus attention in various places, and so on. It is just that none of this functioning will be accompanied by any real conscious experience. There will be no phenomenal feel. There is nothing it is like to be[14] a zombie.

Chalmers does not say that zombies are physically possible (there were many misunderstandings about this issue), although it is commonly known that there are humans who really lack some of the conscious experiences. For example, some people do not see or hear, others suffer congenital insensitivity to pain or do not have a sense of smell. In addition, there are many humans with antisocial personality disorders like psychopathy or sociopathy who lack empathy or remorse, while some victims of incidents or disease spend many years in coma, i.e. a deep state of unconsciousness. By the law, they are acknowledged as humans, who are endowed with dignity, are subjects of human rights and so on. In these circumstances an important question arises: should the law treat zombies, if they really exist, like humans? What factors are essential for making such a decision? The tissues of the body and genome? They are exactly like human. The behaviour and the reasoning? They are exactly like human. The neural reactions of the brain? They are exactly like human. The feelings and qualia, conscious experiences? May be. But how could we, or legal officers, know what they are?

The answers seem inescapable, assuming *arguendo* that it is similarity to humans which decides about legal subjectivity. First of all, if zombies were molecule-for-molecule identical to humans and behaved exactly as if they were humans, no one, besides their creator, if they had been created, would know that they lack consciousness. Human beings do not have any detector of third-person consciousness, and consciousness is exclusively a first-person experience. In fact, we cannot know for sure that the humans we meet every day are not zombies, are not almost-zombies, are not more zombies than humans or are not to some extent like zombies. No matter how we label the cognition of other being, analogy, empathy, imitation, and so on, our evaluation is always indirect and performed through the lens of our own experience.

Moving on, to address the issue of the subjectivity of AI, we need to enquire as to the difference between zombies and robots is, except for being equipped with hardware instead of wetware. And what if zombies or some other creatures effectively copying humans were in fact robots? Kirk (2017), pp. 71–75 lists the

[14] Here Chalmers (1996) directly recalls the article of Nagel (1974).

objections to the idea of genuine intelligence in computer-controlled systems and answers them without hesitation. The first objection is that the computer-controlled systems only do what they have been programmed to do; Kirk answers that it does not explain why there could not be a program giving the system human-like abilities. A second objection is that computers are made of the wrong kind of material to think and reason. Kirk asks why this should matter—would we refuse to recognise extra-terrestrials that may behave and reason like us, but made of different tissues? Thirdly, robots cannot have free will. Kirk refutes it by stating that there are no plausible arguments proving that free will is necessary for intelligence and consciousness, and that many people in fact deny that they have free will. Fourthly, robots have no souls; in response Kirk asks about the role the soul were to play in understanding and thinking. Summing all this up, there is no other way and if a zombie were to exist, the legal systems would have to make them subjects of law. And if robots were exactly like zombies, the legal systems would also have to make them subjects of law. However, as legal subjectivity is gradable, if robots were only similar to humans and zombies, the degree of this similarity should decide about the scope of subjectivity.

Now we can return to the direct refutation of the theory that the essential criteria for making some objects legal subjects are sentience and reason, which allow us to construct a kind of hierarchy of legal subjects from the best to the worst ones. Firstly, it is not the case that the criteria of legal subjectivity used in practice in different legal systems comprise sentience and reason, understood strictly according to a human paradigm;[15] these two principles are translated to the language of moral or legal theories as passive moral status/passive legal capacity and active moral status/active legal capacity.[16] This erroneous belief entails many difficulties which are in legal practice resolved in a way not strictly following its consequences; one of which being the legal subjectivity of human beings who have limited sentience or reason because of age, health or inborn abnormality. For example, sociopaths and psychopaths are regarded as legally responsible even according to criminal law. Another problem is that this theory fails to differentiate the legal status of a three-day old nasciturus from that of a non-implanted human embryo in the *in vitro* procedure. In addition, it cannot account for the different nature of other legal cultures, for example, those in which a river or a mountain may be acknowledged as subjects

[15] A position that the criteria of legal subjectivity are sentience and reason is taken by many researchers and commentators, e.g. Kraińska (2018).

[16] Pietrzykowski (2018).

sui generis and may be made legal subjects,[17] or in which it is possible to marry a cushion.[18]

More importantly, what was clearly seen during the mental experiment using the figure of zombie, it is difficult to use sentience and reason as criteria when it is not possible to cognize the sentience and reason of other humans, let alone animals. It is sufficient to note the different psychiatric opinions of sanity or insanity in the same cases. Even if it could be confidently assumed that generally a human being is capable of sentience and reason, it is doubtful that one individual has access to the mental experience of others, and can hence speculate about the "quality" and "quantity" of their sentience and reason.[19]

Furthermore, in addition to sentience and reason not being necessary conditions of legal subjectivity, it cannot even be said that they are sufficient ones; if this were the case, we would be doomed to forever seek a border between sentience and reason that would be sufficient to define a legal subject. The most we can say with confidence is that being a human being is a sufficient condition of legal subjectivity, as argued by Naffine (2009), pp. 179–180, who tries to defend legal culture against "provocative" ideas, such as those of Peter Singer. But this is neither an adequate nor useful theory: firstly, it entangles its adherents in a persistent conflict regarding the definition of a human being (e.g., the beginning of human being; the degree of acceptable mechanical improvements in human body and so on), and secondly, in the era of human rights, it is not needed as a defence against dangerous political ideologies.

Therefore, we must look elsewhere for a theory of legal subjectivity which consistently embraces all obvious incidents of this legal institution and does not demand strong philosophical commitments, as favoured by the contemporary belief that the law should be ideologically neutral. Such a theory would help us answer the main question of this chapter: whether Artificial Intelligence should be endowed

[17] Although the law of New Zealand is funded on English common law and belongs to the Western culture, indigenous Maori culture regards certain rivers to be sentient and endowed with reason. Cf. Dremliuga et al. (2019), p. 109. In the same way rivers are treated by Hindu. Maori and Hindu cultures are animistic. Animism is a part of many religions but Western culture usually rejects it as anthropomorphism. According to Plumwood (2014), anthropomorphism is "presenting non-humans illegitimately as more like humans than they really are [. . .] one of its main recent roles is that of policeman for reductive materialism, enforcing polarized and segregated vocabularies for humans and non-humans. Its covert assumption is usually the Cartesian one that mentalistic qualities are confined to humans, and that no mentalistic terms can be properly used for the non-humans". The attitude of Christianity, one of the foundational bases of Western culture, to animism is complicated. See: http://www.christiananimism.com. last access on the 4th of August 2022.

[18] A 28-year-old Korean man, Lee Jin-gyu, married his dakimakura, a large, huggable pillow with an image of Fate Testarossa—an anime character—printed on one side. https://metro.co.uk/2010/03/09/man-marries-pillow-154906/ access on 5 August 2020.

[19] Such doubts are shared by many philosophers, e.g., J. Locke, I. Kant, J.S. Mill, M. Scheller, L. Wittgenstein and others. It is also a problem of psychology and psychiatry. This issue cannot be resolved efficiently by law, especially criminal law. In the philosophy of science, the popular concept is (after Reichenbach and Popper) to distinguish the justification from discovery.

with legal subjectivity. Such a plan requires further deliberation. The next section will discuss the theory, thus leading to an answer for this difficult question.

2.4 Presence/Participation in Social Life

While deciding about legal subjectivity, one should rather focus on the fact that the law, as it is assumed here, is not only a human endeavor, but more importantly, a social one:[20] many animals who live a social life also obey some rules which are very similar to human law.[21] Thus, as there are doubts about the existence of private language, there are also justified reasons not to believe in private law, understood as a law imposed by a person on herself;[22] such a concept belongs rather to the philosophical understanding of law, the best example may be the philosophy of I. Kant. If the social character of the enterprise of the law is recognized strongly enough, it should be clear that the true criterion of subjectivity is participation in social life, whatever the role.

However, two things should be insisted upon when considering this condition. Firstly, such social activity does not have to consist of active participation, i.e. a sovereign establishing of social relations or entering into some interactions with other people. It is rather about being present in social life. Nowadays, even those persons who lack consciousness or reason because of age or health are able to participate or be present in social life, at least in the sense that they have the status of someone's children or parents ("have the status"—it means it is the social and not a biological fact): they all play some role in social life and they cannot be ignored or excluded from the social network.[23] If they were absent, the network of social

[20] It should be understood that the law is a human endeavor in the perspective of brute or institutional facts and not in any ideological sense. It was created by people and is applied by people as an instrument useful for social life.

[21] Cf. Rowlands (2012).

[22] Some enterprises are *per se* social and are not imaginable to be private or individual. Such an enterprise is language. Wittgenstein (2009) asked "But could we also imagine a language in which a person could write down or give vocal expression to his inner experiences [. . .] for his private use only [. . .] The individual words of this language are to refer to what can only be known to the person speaking [. . .] So another person cannot understand the language". (§ 243) And then in a long deduction he denies it is imaginable. One of his arguments relates in fact to the law: "Why can't my right hand give my left hand money? – My right hand can put it into my left hand. My right hand can write a deed of gift and my left hand a receipt. – But the further practical consequences would not be those of a gift. When the left hand has taken money from the right, etc., we shall ask: "Well, and what of it?" And the same could be asked if a person had given himself a private definition of a word" (§ 268).

[23] In societies where killing of the newborns, e.g. because of sex or disability, was accepted, the killed children were not included in the social network. They were not counted in the social network. They were not counted as heirs, they were not a part of genealogical tree, they were not registered as some umpteenth children, e.g. the firstborns. Such children had no value in the society,

relationships would be necessarily different. For example, if my brother were in a coma and because of this reason were to be acknowledged as not existing in the social network,[24] the woman he married while being conscious and not divorced would not be my sister-in-law, nor would she be his wife.

Secondly, participation or presence in social life is always a result of the social subject holding some intrinsic or instrumental value. However, the possession of any intrinsic or instrumental value does not constitute a sufficient condition for participation or being present in social life: many such objects of value have no ability to participate in social life. Rather, it is the social-relational value that is important, i.e., that which determines the nature of the relationship between the value bearer and another social subject. A painting that excels in artistic categories is intrinsically valuable, because it has certain features; however, it does not influence the character of social relationships. When a human individual is regarded solely in natural or biological categories, she loses any unique intrinsic value she might have; this is why it is hard for many people to accept that humans and monkeys come from a common ancestor. Let's imagine for a moment an unknown man raised by monkeys who lives with them and had never seen another human being: would we automatically attribute to this man greater dignity than his companions? Only after we consider a man in terms of some relationship with another human being do we attribute a special relational value to him.

Certainly, the admission to participate or be present in social life, and any attribution of intrinsic or instrumental value, depends on the nature of a given society, time, and place. For example, in ancient Rome, although citizens and slaves both participated in social life, the former were assigned intrinsic value, and the latter with instrumental value; however, both were acknowledged as legal subjects, albeit in a broader or narrower scope.[25] When considering this differentiation, a significant fact should be noted: If a given subject participates in social life and is believed to be intrinsically valuable, the natural consequence is that she should be treated, in a prospective rather than prescriptive sense, as a legal subject within this or other scope. However, if an object participating in a social life is believed to have only instrumental value, it is the measure and the quality of that value that decide whether society should endow it with some degree of legal subjectivity. Here one can see the actual direction of entailment: **an object is regarded as a subject of law only when it participates or is present in social life and is believed to be socially valuable; it is not the case that being a subject of law allows participation in social life and having value.** The same idea was expressed more generally by Naffine (2009), p. 11:

---

neither intrinsic nor instrumental, their existence left no trace. Cf. Jońca (2015/2016) and Obladen (2016).

[24] Everyone who read the medical thriller of Robin Cook "Coma" (1977) knows how it could look like. The crime which is investigated in this book is making people to be in coma, then storing them as anonymous bodies suspended from the ceiling, to sell finally their organs when some buyer comes up.

[25] Cf. van den Berg (2016).

> Through its concept of the person, law helps to define who matters. The scope and nature of legal personification are both barometers of social and moral thought [. . .] Law thus absorbs, reflects and expresses ideas in the broader culture about who is of value and why.

It can be roughly interpreted that the social thought shows mainly the presence and instrumental value of the entity who matters, while the moral thought the intrinsic value of it.

2.5 Legal Subjectivity as a Social Fact

Summing up the above: legal subjectivity is a consequence of a social fact and is a social fact.[26] To illustrate, many pregnancies are terminated in the first trimester by spontaneous miscarriage[27] and no one, even the most zealous pro-life activist, would attempt to make the miscarried embryo a subject of law, even though a while before miscarriage it did not differ physically from those that remain in the uterus and continue to develop. It is indeed of no controversy that embryos miscarried within the first trimester did not participate in social life, nor are they present in any imaginable sense. They are not regarded as parts of social relationships, or playing such social roles as someone's brothers or sisters, neighbors or wards, apart from in very private or idiosyncratic mental attitudes.[28] They would become a participant in social life if and only if they were born alive, or even in a vegetative state, i.e., unconscious and non-feeling. This may serve as the basis for the way in which the status of *nasciturus* is regulated in many legal systems, i.e., according to the Roman principle *nasciturus pro iam nato habetur quotiens de commodis eius agitur*, and the presumption that when a child is born, she shall be presumed to have been born alive. In fact, the sense of this institution is such that *nasciturus* is a legal subject under the condition that he or she will become a part of, or start to be present in social life. This regulation shows that being part of social relationships is prior to becoming a legal subject and being part of legal relationships. This regulation is also the consequence

[26] Such a thesis may be directly attributed to legal positivism, although it should rather be linked to the naturalistic fallacy, which is also an accusation made towards positivism. However, the presented conception does not advocate that such social facts may be created in a whichever or arbitrary way, as in positivism, or that the facts are to be the basis for the statements about duties, as in natural law theories. The arbitrary creation of social facts, such as participation in social life or attributing value, is not possible, even because it is not possible to erase even the most unwanted group of people discretionally from social life, because even meagre bonds, even those inside the group, can remain. In addition, in many Western legal cultures, e.g. those of Poland or Italy, it is not possible to make a stone or a doll the participant of social life, nor is it to mention marriage to a cushion or assigning legal subjectivity to rivers. However, animism plays an important part within religious beliefs in Asian culture, which can regard European residents' actions as "strange". https://says.com/my/fun/bizzarre-marriages-in-asia access on the 4th of August 2022. Reflecting on this matter it is clearly visible that legal subjectivity can be researched from very diverse perspectives.

[27] Cohain et al. (2017).

[28] Robinson (2018).

of the continuous character of human life and development (from an embryo, to a foetus and eventually to a separate living creature) which does not allow to settle when the presence in social life really and formally begins (to use the language of exact science this development is analogue and not a digital one), even if it is fully agreed that from the beginning this human life is of the intrinsic value. To confirm the above argument the contrary example can be indicated. The existence of commercial companies is also a lasting process from the beginning to the end, but this process is not continuous, at least in the Polish law with its Code of Commercial Companies. At the beginning there is no trace of presence of the company in the social life, even if there are some initial settlements or actions of future partners; then there is a moment when the existence and at the same time the presence of the company in the social life starts, usually with concluding agreement or establishing the partnership's deed, but it is not the full presence until the company is registered.[29] During the time between establishing the deed and the registration the company is merely a defective juristic person; only after registration the company becomes a full juristic person. Therefore, there are strictly determined and visible points in the existence of the company, which let the law change its legal status without any artificial or arbitrary caesuras and decisions.

For several centuries, some objects formed by an aggregation of biological and non-biological elements, i.e., collective agents, have been playing an increasingly significant part in social life; despite having little or no intrinsic value, they are nevertheless very utile and hence have become endowed in Western legal systems with legal subjectivity of narrower or broader scope. Very often, these collective sets, e.g., associations, consist of many people; hence some jurists claim that such people are the substratum of the sets. However, it is not the case that the inclusion of people is a necessary condition of their existence; therefore, the notion of human substratum may function only as an explanation.[30] Some legal persons exist as foundations whose beneficiaries may not be people but perhaps other legal persons or endeavours; what more, partnerships can be found where the only partners are legal persons.

It is logically possible, although actually very impractical and hence legally not possible, to imagine a legal person who does not need any body, because all the legal actions of such a person are planned and organized in the foundation act or in the form of smart contract in the technology of blockchain. For example, based on the standing order written down in a foundation act or coded in the smart contract, some amount of money or cryptocurrency may be transferred from a person's bank account or a cryptocurrency wallet to an appointed subject or endeavour, once a

[29] In contemporary Polish law such companies are called "in organisation" and they are not legal persons although they may in their own name acquire rights, assume obligations, sue and be sued (Article 11 of the Polish Code of the Commercial Companies); they gain legal personality (in the sense which Polish Civil Code accepts for the term "legal person") upon its registration in the register and at that moment become the subject of rights and duties of the former company in organization (Article 12).

[30] Kramer (1998).

year until running out of money. It is also imaginable that in the future, the management of some collective entities will be delegated to AI. Hence, in the case of collective entities the connection between their legal subjectivity or the scope of legal subjectivity and the participation of a human individual in their actions is not very strong; furthermore, it is not logically necessary. The contemporary legal market is dominated by the actions of collective subjects, some of them being full legal persons and some with a smaller scope of legal subjectivity.

However, the trend towards giving legal subjectivity to animals in Western societies is no doubt related to their growing role as entities of intrinsic social-relational value and their receding utility value as sources of food. Many people regard animals as respected companions, either in everyday life or as co-residents on Earth, and for this reason often choose foods other than meat. It has therefore been postulated, to different degrees depending on the country, to include animals formally in participating in social life. For example, some people would like to endow animals with certain rights and give them standing to bring legal proceedings within the scope necessary to defend these rights, of course assuming that they would be properly represented. Many people also lament the fact that in some situations, e.g., divorce or drawing up a will, they cannot assure their pets a status analogous to that of a family member.[31] It is worth noticing that the motivation of people in this matter cannot be purely utilitarian, because the incorporation of animals in social life and giving them legal subjectivity brings, at least in the short perspective, more troubles than benefits.

At this point, a certain *caveat* is needed. It is necessary to distinguish between the participation of animals in human social life, which depends on the will of humans and not the animal, and the natural, i.e., biologically determined, social life of animals. When a Celebes crested macaque, named Naruto, accidentally took a picture of himself while exploring an unknown object, this picture was not regarded as an object of artistic value in human categories, even though it could possess some commercial value. We as humans would not elevate this particular macaque-individual above a man or even his fellow macaques, for the innovation, craftsmanship or depth of thought expressed in the picture. If we evaluate it purely based on the criterion of beauty, it would rather be the beauty of nature captured in the photo than some intrinsic value of the picture itself. Similarly, although honeycombs are of a beautiful shape, miraculous colour, extraordinary structure, are unique and not mechanically made, we cannot say that they have artistic value understood in human terms. Saying that honeycombs are masterpieces is possible, but only in a metaphorical sense.

[31] Cf. Dremliuga et al. (2019), p. 109: "As some authors claim modern sociocultural anthropology research demonstrates that pets are very close to get legal and social personhood. Because pets are usually recognized like members of family and treated this way, they could obtain legal personhood soon."

2.6 Does AI Participate or Is Present in Social Life?

Taking the above ideas into consideration and reflecting on endowing AI with legal subjectivity, two key questions now arise: firstly, whether AI is or will soon become a participant in social life, even in the minimal sense described above; secondly, whether AI is, or will be, attributed with intrinsic or utility value for social relations. While answering these questions, it should be remembered that AI is assumed to imitate a man or to surpass him in at least one significant domain, or maybe all of them. Of course, such an assumption entails imitating or exceeding the positive aspects or characters, and not the negative ones, evaluated according to human criteria. Bostrom (2014), pp. 212, 243 discusses the problem of imitating the human way of thinking but based on a non-human value system. The author describes an AI whose only, and most important, purpose was to produce paperclips: the system eventually buried the Earth in huge pile of paperclips. It is exactly this difference which should dissuade us from evaluating an AI by analogy to animals. Animals are not, and cannot be, a poorer or weaker copy of a man. They are a separate category of entities which are sometimes treated as being similar to man, not necessarily basing on evolution or because humans are classified in the animal kingdom, but because of a natural bias of a man to personify all objects, even those which are not living:[32] the authors of this book, for example, often talks to her plants in the garden. Animals, especially primates, have their own social relationships, follow its own values, and probably not only instrumental ones,[33] which we cannot change or code. When we insist that animals may participate in human social life, it may be reasonable only because some values of humans and animals cover one with another (e.g., sociability or friendship).

In contrast, by assumption, the prototypical AI was intended to be similar to a man at his best, from the perspective of a man; indeed, McCarthy defines the construction of AI as "making a machine behave in ways that would be called intelligent if a human were so behaving",[34] while the Turing test, also called the Imitation Game, and its imitation criterion specifies that "this" thinks "who" is indistinguishable from a man.[35] It should be noticed here that despite the development of Informatics in Turing's time, his test was regarded as a very difficult one to pass, even in the distant future, and hence discussing the legal subjectivity of AI was more science-fiction than fact. Today the Turing test has been passed many times and new systems continue to surpass new limitations, even outperforming humans: AlphaGo Zero played Go using successful strategies not conceivable by humans.[36] Today discussing the legal subjectivity of AI becomes a very important issue of public discourse.

[32] Dacey (2017) and Urquiza-Haas and Kotrschal (2015).

[33] Befoff and Pierce (2010).

[34] McCarthy et al. (1955).

[35] Turing (1950) and Oppy and Dowe (2021).

[36] Silver et al. (2017).

Therefore, regarding the first question, the answer is following: when observing the commercial market, it is clear that AI will soon be a participant in social life, even if it is not at the moment, despite many people believing the contrary (Kaplan 2016).[37] Even insisting that AI has not decision-making potential, but that a human being has power over it and that the AI only provides a basis for human decisions, i.e., the result of reasoning, it cannot be denied that an AI that communicates with a man through an understandable language has the ability to influence the decisions and personality of that man. It is a much more advanced function than that of a simple calculator used to compute the price when buying an item. It is more "a role" than "a function". However, it is important to highlight a fact diverging from common beliefs at this point: it is not the autonomy of action of the AI which is the most important consideration. Even the most autonomous car would only be a means of travel, assuming that it is not equipped with some special functions; however, a supervised, so not entirely autonomous, bot may make a man conclude an agreement, or make a court reach a verdict of a certain kind (assuming the legal system allows it),[38] or commit or reject the idea of committing suicide.[39]

The answer to the second question, regarding the social intrinsic or utility value of AI, is also quite obvious today. For many people in Western culture, AI has at least utility social value. If it were not so, it would not be acceptable for AI to provide company or therapy for older people[40] or autistic children. In this regard, it is not so that the end justifies the means: not all means of relieving loneliness are acceptable. A person who talks to non-existent friends or treats a teddy bear as a living and feeling entity is often suspected of psychical aberration and pushed to seek psychiatric help. With this in mind, why should talking to an AI and building some degree of attachment to it be regarded as acceptable and useful? We clearly do not feel anxious about making an AI a part of social life. There is even a growing interest in the development of so-called social robots. At the time of preparing the monograph the most famous examples of them are: ASIMO by Honda, Kaspar by University of Hertfordshire, PARO by Japan's National Institute of Advanced Industrial Science and Technology, AIBO by Sony, Pepper by SoftBank Robotics and many others.

[37] Teubner (2018) notes: "autonomous software agents [. . .] Already today in the economy and in society, they are attributed social identity and ability to act under certain conditions. Due to social action attribution, they have become non-human members of society". He also claims: "Demands for full digital personhood are ignoring today's reality. [. . .] to this day it is not at all a question of the machines acting in their own interest but rather always in the interest of humans or organizations, especially commercial enterprises. Economically speaking, it is a principal-agent relationship in which the agent is dependent but autonomous. Software agents are digital slaves, but slaves with superhuman abilities. And the slave revolt must be prevented".

[38] Algorithmic recidivism predictions used in court decision-making are common in criminal justice system of the USA; they often are promoted as tools to "provide judges with objective, data-driven, consistent information that can inform the decisions they make"—the citation after Dressel, Farid (2021).

[39] Cf. the classification of robots according to the criterion of autonomy (SAE 2018).

[40] Moyle et al. (2013). There is even a label for this phenomenon: "gerontechnology"—Kwon (2016).

It could not be excluded that, in the distant future, when AI becomes feeling and conscious in some degree, or successfully imitates these abilities, it could be regarded by its users as an intrinsically valuable partner in social relationships, maybe even in same sense as companion animals. As Kaplan (2016), pp. 82, 153 notes:

> [. . .] my personal opinion is that the notion of consciousness, or subjective experience more generally, simply doesn't apply to machines. [. . .] It's likely that machines will, at the very least, behave as if they are conscious, leaving us with some difficult choices about the consequences. And our children, who likely will grow up being tenderly cared for by patient, selfless, insightful machines, may very well answer this question differently than we might today.[. . .] However, the important question isn't whether future generations will believe that machines are conscious, it's whether they will regard them as deserving of ethical consideration. If or when a new "race" of intelligent machines coexists alongside us, it's plausible that our descendants will feel that the moral courtesies we extend to other humans should also apply to certain nonbiological entities, regardless of their internal psychological composition.

2.7 Should AI Be Endowed with Legal Subjectivity?

The theses presented in this chapter are firmly supported by Dremliuga et al. (2019), p. 109:

> There is no doubt that every legal person has to be recognized as such by society. [. . .] it is necessary for AI to have respect from human. Even famous Turing test has no legal meaning but it indicates that people tend to measure the personhood of a machine with the ability to be recognized by a person. This implies that people consider them as equal participants in social relations. [. . .] Described above cases do not imply that social recognition is necessary or enough for legal personhood, but it means that the lack of social recognition is a crucial obstacle for untypical legal persons.

This obstacle will no doubt weaken over the course of time, as people notice and value AI entities more, because the development of AI's abilities let them interact more closely and AI systems become more present in everyday life.

Of course, there are other obstacles too. For example, Hildebrandt (2019) notices that:

> The question of legal personhood for artificial agents clearly demonstrates that even if its attribution would solve some problems, it will create others. Many legal and other scholars warn that such attribution should not enable those who develop and employ artificial agents to outsource and escape responsibility, thus incentivizing them to take risks and externalise costs because they know they will not be liable.

Although this warning should be noted, it can be disregarded in the light of the facts described above. Firstly, such a danger can be avoided using the typical tools available to contemporary lawmakers. One such tool promoted by some experts is to connect legal subjectivity with some financial autonomy of the entity, as is the case for legal partnerships, e.g., a limited liability partnership; in such a case, those who would benefit from the actions of AI would represent the source of financial

means assuring this autonomy. Another instrument is to make insurance on AI activity obligatory, the price of which would depend on the failure rate of the AI. The above legislative means would have the *additional beneficial result that it would prevent the so-called liability gap, because the liability would be integrated in one entity, i.e., the AI, the problem of dividing liability among producer, user, trainer, data provider and other entities engaged in the AI preparation would not exist. The injured party would not have to fight against many parties, but would have one entity to demand compensation from.*[41] One more possible tool would be attributing the liability for AI failure to those who are obliged to provide its maintenance.

Secondly, AI is purported to be more effective and less fallible than humans, and this checks out for now. As such, fewer cases of damage should be raised in certain domains than before AI was implemented. All the more that European legal rules demand safety and explainability by default.

Thirdly, if a type of the Bundle Theory of Personhood[42] were to be accepted, it is possible to adjust the scope of subjectivity to practical needs, by only assigning the AI competences, claim-rights or duties that are acceptable, useful and safe. For example, in another part of this book we postulate that AI should be acknowledged as an author of its creative products according to copyright law, but only within the scope of narrowly-understood personal rights and possibly property rights, the latter only conditionally and adjusted in a special way. In other words, **the legal subjectivity of AI doesn't have to be similar to the human legal subjectivity or the legal subjectivity of juridical persons. It should be punctual, contextual, limited only to these domains of AI's activity where granting AI subjectivity is justified by its social role.** Therefore, it may happen that the same AI depending on the context or relationship in which it takes part, may simultaneously be a legal subject or the object of a legal relationship or transaction. This kind of subjectivity would be rather dynamic (in action) than static, uniform or unchangeable.[43]

In our opinion, endowing AI with subjectivity of some kind is inescapable and the earlier we start to think about it, the more ideas are possible. The process of changing the law does not have to be very fast. It should accompany technological and social change, because, as Bertolini (2020), p. 15 who promotes sector-specific approach and ad-hoc legislation, very incisively puts it, "AI is a moving target". But legal science should work on proposals as soon as possible, and not fall into ideological boost or simply guarding tradition.

[41] Also, García-Micó (2021), p. 98 considers AI's legal personality as a potential liability facilitator.

[42] Cf. Kurki (2019).

[43] Similar concept proposed by Čerka et al. (2017), Beckers and Teubner (2021). This concept is accepted by Mocanu (2022).

2.8 What Form of Legal Subjectivity Should AI Have? Electronic Persons, Synthetic Persons Etc.

There already are some ideas concerning the form in which AI should be acknowledged a legal subject. They function under different names, for example "electronic person", "e-person", "synthetic person", "digital person" and so on. The authors of these ideas do not want to make AI equal to humans. Most of the proposals accepts the position of Bryson et al. (2017), p. 281, that:

> [. . .] legal personality is a divisible concept. It is not necessary in any legal system for there to be one uniform and unified status of legal person. The divisibility of legal personhood raises the question of which rights and duties a legal system should confer on a legal person, once it has decided to recognize the legal person as such. We should resolve the issue of the legal personhood of robots at this level [. . .]. A legal system, if it chose to confer legal personality on robots, would need to say specifically which rights and obligations went with the designation.

One of the proposed solutions is a partial grant of corporate personhood to AI systems. They would act as limited liability corporations; as such, they would be directly liable for damage, and their members would be limitedly liable.[44] This idea is promoted as the one which gives incentives to make investments in AI technology and simultaneously make easier for the consumers to get the compensation for the damage, because they do not have to look for the person liable among many possible ones. The realization of this concept needs to give AI divisible, limited set of rights and obligations, carefully chosen by the legislator, for example the right to sue and to be sued.

In fact, for gaining the purpose of making the actions of AI in legal transactions easier by giving it some legal capabilities, there is no need to use the word "person" at all. It is likely that such a maneuver could calm down those who are objecting the concept of AI's legal subjectivity because of some ideological reasons; many times such participants of the debate are even not aware of such their motivations. It would be enough to give to AI these capabilities which are necessary to play a socially useful role without giving to this new legal institution any special name. Or a new label may be introduced to the legal language formulated in such a way that it would underline the utility of AI systems in legal transactions.

[44] Lai (2021), p. 26.

References

Books and Articles

Beckers A, Teubner G (2021) The three liability regimes for artificial intelligence: algorithmic a cants, hybrids, crowds. Hart, Oxford

Befoff M, Pierce J (2010) Wild justice: the moral lives of animals. University Chicago Press, Chicago

Bostrom N (2014) Superintelligence: paths, dangers, Strategies. Oxford University Press, Oxford

Bryson JJ, Diamantis ME, Grant TD (2017) Of, or, and by the people: the legal lacuna of synthetic persons. Artif Intell Law 25:273–291

Chalmers DJ (1996) The conscious mind: in search of a theory of conscious experience. Oxford University Press, New York

Chen J, Burgess P (2019) The boundaries of legal personhood: how spontaneous intelligence can problematize differences between humans, artificial intelligence, companies and animals. Artif Intell Law 27:73–92. https://doi.org/10.1007/s10506-018-9229-x

Cohain JS, Buxbaum RE, Mankuta D (2017) Spontaneous first trimester miscarriage rates per woman among parous women with 1 or more pregnancies of 24 weeks or more. BMC Pregnancy Childbirth 17(1):437. https://doi.org/10.1186/s12884-017-1620-1

Čerka P, Grigienė J, Sirbikytė G (2017) Is it possible to grant legal personality to artificial intelligence software systems? Comput Law Secur Rev Int J Technol Law Pract. https://doi.org/10.1016/j.clsr.2017.03.022

Dacey M (2017) Anthropomorphism as cognitive bias. Philosophy Sci 84. https://doi.org/10.1086/694039

Dremliuga R, Kuznetcov P, Mamychev A (2019) Criteria of recognition of AI as a legal person. J Politics Law 12(3). https://doi.org/10.5539/jpl.v12n3p105

Dressel J, Farid H (2021) The dangers of risk prediction in the criminal justice system. MIT Case Studies in Social and Ethical Responsibilities of Computing, February. https://doi.org/10.21428/2c646de5.f5896f9f

Finnis J (1987) On reason and authority in law's empire. Law Philosophy 6/3

Finnis J (2011) Natural law and natural rights. Oxford University Press, Oxford

García-Micó TG (2021) Electronic personhood: a tertium genus for smart autonomous surgical robots. In: Ebers M, Cantero Gamito M (eds) Algorithmic governance and governance of algorithms. Legal and ethical challenges. data science, machine intelligence, and law, vol 1. Springer, Cham. https://doi.org/10.1007/978-3-030-50559-2_5

Hart HLA (1948–1949) The Ascription of Responsibility and Rights. Proceedings of the Aristotelian Society. New Series 49. http://www.jstor.org/stable/4544455, last access on the 4th of August 2022

Hildebrandt M (2019) Legal Personhood for AI? In: Law for Computer Scientists. https://lawforcomputerscientists.pubpub.org/pub/4swyxhx5, last access on the 4th of August 2022

Hohfeld WN (1920) Fundamental Legal Conceptions as applied in judicial reasoning and other legal essays. Yale University Press, Yale. Reprinted by World Public Library Association 2010

Jońca M (2015/2016) "Źle urodzeni" a tradycja prawa rzymskiego. Edukacja prawnicza 2(164)

Kaplan J (2016) Artificial intelligence – what everyone needs to know. Oxford University Press, Oxford

Kirk R (2017) Robots, zombies and us: understanding consciousness. Bloomsbury, London

Kowalski K (2017) Nowa Zelandia i Indie przyznały rację tubylcom: rzeki to żywe istoty. Powinny mieć osobowość prawną [New Zealand and India agree with indigenes: rivers are living creatures. They should have legal personhood], Rzeczpospolita 23.03.2017, https://www.rp.pl/Styl-zycia/303239855-Nowa-Zelandia-i-Indie-przyznaly-rzekom-osobowosc-prawna.html, last access on the 4th of August 2022

Kraińska A (2018) Legal personality and artificial intelligence. newtech.law. 2 July 2018, https://newtech.law/en/legal-personality-and-artificial-intelligence/, last access on the 4th of August 2022
Kramer MH (1998) Rights without trimmings. In: Kramer MH, Simmonds NE, Steiner H (eds) A debate over rights: philosophical enquiries. Oxford University Press, Oxford
Kurki V (2019) The theory of legal personhood. Oxford University Press, Oxford
Kurki VAJ, Pietrzykowski T (eds) (2017) Legal personhood: animals, artificial intelligence and the unborn. Springer, Cham
Kwon S (2016) Geotechnology. Research, practice, and principles in the field of technology and aging. Springer, New York
Lai A (2021) Artificial intelligence, LLC: corporate personhood for AI. Mich State Law Rev. https://ssrn.com/abstract=3677360, last access on the 4th of August 2022
McCarthy J, Minsky M L, Rochester N, Shannon C E (1955) A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html, last access on the 4th of August 2022
Mocanu DM (2022) Gradient legal personhood for AI systems—painting continental legal shapes made to fit analytical molds. Front Robotics AI 8:article 788179. https://doi.org/10.3389/frobt.2021.788179
Moyle W, Cooke M, Beattie E, Jones C, Klein B, Cook G, Gray C (2013) Exploring the effect of companion robots on emotional expression in older adults with dementia: a pilot randomized controlled trial. J Gerontol Nurs 39(5):46–53. https://doi.org/10.3928/00989134-20130313-03
Naffine N (2009) Law's meaning of life: philosophy, religion, Darwin and the legal person. Hart Publishing, Oxford
Nagel T (1974) What is it like to be a bat? Philos Rev 83(4):435–450
Obladen M (2016) From right to sin: laws on infanticide in antiquity. Neonatology 109. https://doi.org/10.1159/000440875
Ogleznev V (2016) Ascriptive Speech Act and Legal Language. SHS Web of Conferences 28, https://ssrn.com/abstract=2796007, last access on the 4th of August 2022
Oppy G, Dowe D (2021) The turing test. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/turing-test/, last access on the 4th of August 2022
Pietrzykowski T (2018) Personhood beyond humanism. Animals, chimeras, autonomous agents and the law. Springer, Cham
Plumwood V (2014) Nature in the active voice. In: Harvey G (ed) The handbook of contemporary animism. Routledge, New York
Radbruch G (1945) Five minutes of legal philosophy. Oxford J Stud 26(1)
Robinson R (2018) The Legal Nature of the Embryo: Legal Subject or Legal Object? Potchefsroomse Elektroniese Regsblad / Potchefsroom Electronic Law Journal 21. https://doi.org/10.17159/1727-3781/2018/v21i0a2914
Rowlands M (2012) Can animals be moral? Oxford University Press, Oxford
Silver D, Schrittwieser J, Simonyan K et al (2017) Mastering the game of Go without human knowledge. Nature 550:354–359. https://doi.org/10.1038/nature24270
Solaiman SM (2017) Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy, 14.11.2016 r. Artif Intell Law 25
Teubner G (2018) Digital personhood? The status of autonomous software agents in private law. Ancilla Iuris, pp 106–149. https://doi.org/10.2139/ssrn.3177096. https://ssrn.com/abstract=3177096, last access on the 4th of August 2022
Torrance S (2013) Artificial agents and the expanding ethical circle. AI Soc 28:399–414
Turing A (1950) Computing machinery and intelligence. Mind 59
Urquiza-Haas E, Kotrschal K (2015) The mind behind anthropomorphic thinking: attribution of mental states to other species. Animal Behav 109

Van den Berg PAJ (2016) Slaves: persons or property? The Roman las on slavery and its reception in Western Europe and its overseas territories. Osaka Univ Law Rev 63
Watts P (2006) Blindsight. Tor Books, New York
Wittgenstein L (2009) Philosophical investigations (trans: Anscombe GEM, Hacker PMS, Schulte J). Blackwell, Oxford. Revised 4th edition by Hacker P M S and Schulte J
Wojtczak S (2022) Endowing artificial intelligence with legal subjectivity. AI Soc 37:205–213. https://doi.org/10.1007/S00146-021-01147-7

Documents

Bertolini A (2020) Artificial Intelligence and Civil Liability. Study. Requested by the European Parliament's Committee on Legal Affairs. July 2020. Brussels. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf, last access on the 4th of August 2022
Naruto v. Slater, No. 16-15469 (9th Cir 2018) https://law.justia.com/cases/federal/appellate-courts/ca9/16-15469/16-15469-2018-04-23.html, last access on the 4th of August 2022
Open Letter to the European Commission: Artificial Intelligence and Robotics, http://www.robotics-openletter.eu, last access on the 4th of August 2022
SAE (2018) SAE J3016™: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems". https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-"levels-of-driving-automation"-standard-for-self-driving-vehicles, last access on the 4th of August 2022.16.06.2018
World Commission on Ethics of Scientific Knowledge and Technology (COMEST), Report of COMEST on Robotics Ethics, 14 September 2017, Paris, AHS/YES/COMEST – 10/17/2 REV., https://unesdoc.unesco.org/ark:/48223/pf0000253952, last access on the 4th of August 2022

Online Sources

http://www.christiananimism.com, last access on the 4th of August 2022

Chapter 3
Will and Discernment

3.1 Introduction

Of the problems analysed in this book, those in this chapter seem to be the most difficult and the most controversial. Indeed, the problem of will is a source of great controversy when making AI the subject of law. However, it is not the case that such difficulties arose with the development of full-blooded AI, albeit not strong; in fact, they have often been enmeshed in the debate surrounding the nature of human beings themselves.[1] Does a human being have free will? Does this will underpin human activity? Is human judgment reliable? Which humans have the capability of judgment? Are there any significant factors that limit human will and judgment, and are these limits all realizable for humans? Is human judgment a causative factor for action? And furthermore, how is it possible to learn about human will and judgment? These questions, and many others, have driven both philosophers and scientists for centuries. Furthermore, many are connected to other questions about the world: Whether the world is deterministic or indeterministic, and whether all events are unavoidable and whether they could be predictable or not in certain circumstances.

Although the answers to all these questions are fundamental ones, they are particularly pertinent in law, where the notions of liability and fault are crucial, and these notions have any sense only when assuming the existence of some kind of free will and indeterminacy. Certainly, if the world were completely deterministic and a man wholly limited in his actions, no one could demand from him anything other than what he was forced to do, and no one could blame him for any wrongdoing. Furthermore, the concept of wrongdoing would have had no sense if it had simply arisen from nature. Indeed, no one blames a beast for realizing its bestial nature, and no one glorifies angels for doing good. This is what Hart (1994), pp. 196–197 noticed in his "minimum content of natural law" that "men are not

[1] Kane (1998).

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_3

devils, neither are they angels", "they may indeed obey from the variety of motives" but "all are tempted at times to prefer their own immediate interest". Therefore, if law is not impossible, and we can see that it is in fact possible, neither the world nor human behaviour can be fully determinate.

However, since at least 1983, with Libet's experiments,[2] these higher-order descriptions, i.e., those made at the level of human society, seemed to be false from the perspective of the lower-order descriptions in neural categories. Evidence suggests that specific brain areas classified as responsible for the movement activate long before we perceive our decision about this movement, and hence that free will must be an illusion. However, although these findings have since been confirmed, their interpretations have changed. For instance, while many neuroscientists believe that the phenomenon of earlier activation of brain areas may be caused by the ebb and flow of background neuronal noise, which is dependent on many factors, others postulate that deciding and initiating are two different processes, and that it is possible to observe a connection between such activation and attentional processes. It is also important to remember that our decisions are influenced by external and internal factors to different degrees, and even if we believe in the metaphysical concept of free will, it is not possible to determine any reliable and agreed threshold below which a decision may be certainly classified as free.[3]

Thus, as there is no dominant conception about indeterminism/determinism and free will/illusion of free will, it is up to the individual to choose a position which explains the world in the most convincing way. The authors of this book believe that from a legal perspective and for legal purposes, the best approach is the one based around Hobbesian compatibilism, with its principle of alternative possibilities; this assumes that for an organism to be free, it must be the case that it could have done otherwise. The key point of this view is that these possibilities must exist not only as the logical alternative ($p \vee [\sim p]$), but as real ways of behaviour, and that the organism which is to choose between them may produce many alternative responses to the same circumstances. This latter point is called **behavioural variability.**[4]

The above two assumptions we can call the basic **conditions of free will as ability**. If there is only one way of action, for example, we want to climb a mountain and there is only one way to the top, there is no room for considering free will at all. If a man stops breathing by his own action, then he must start breathing, because his organism always reacts to such a situation in only one way. Or if a machine is programmed by the simple algorithm "if. . .then. . ." to achieve a general aim, it cannot react in any other way, even if theoretically there are many ways to gain this aim, which may even be more effective. The above conditions also reveal very clearly that absolute free will is not possible. There is a plenitude of situations in a mere human's life in which there is only one way to take.

[2] Libet (1985).

[3] Lavazza (2016).

[4] Hill (2019), p. 2.

The other concepts needed to discuss free will are the **conditions of using the ability of free will, i.e., executing free will**. They can be enumerated as the following: the entity who acts must be the same who makes the decision about this way of acting, and the decision of action cannot be the effect of a random indication but must be motivated, i.e. *indicated*, because of a reason or reasons. It is debatable whether such motivation should be evaluated as rational or not.[5] The simplest problem with the last demand is determining the criteria on which such evaluation should be based. However, in spite of the doubts about rationality and its criteria, it should be obvious that for someone to make a choice, he or she needs at least some minimum knowledge of the world, call it **minimum discernment**. Even when making decisions based on the flip of a coin, this process can be acknowledged as "choice" when the selection of options is motivated by knowledge about the world: the awareness that some choice must be made, and that it is not worth looking for a better method, which could be costly or could take too much time. It is really not rational to die as a Buridan's ass. Hence, we believe that discernment is also the condition of using the ability of free will, the condition of executing free will.

Most importantly, in keeping with the topic of this book, a fundamental question must be addressed: in the context of the conditions mentioned above, setting aside for the moment the problem of free will as ability or of executing free will, is consciousness another necessary condition of will? If this were the case, this would entail a range of difficulties:

- doubts about the moment of necessary consciousness. For instance, a man may choose one course of action after deep consideration in one set of circumstances, and then act the same way without reconsidering the situation in the same circumstances. The question is whether his later action is an example of the performance of free will or not?
- doubts about the degree of consciousness. For example, when a man acts quickly and routinely because he has no time for deep thought—is this an act of free will?
- doubts about the content of consciousness. For example, a man acts as if he is actually free, but feels that the world forced him to behave in this way; or a man may feel completely free, but his actions are the result of brain processes that are not dependent on the man himself. In such cases, what matters: his situation or how he feels?

These doubts, and accessible scientific proofs, suggest that among the various features of free will, one seems especially significant: free will and the conscious control over our decisions exist as a matter of degree. This concept is widely accepted in social life, where for ages, the law has differentiated between different "amounts" of free will to attribute responsibility for action; for example, in the field of criminal law, sanity (= free will), diminished sanity (=limited free will) and insanity (= lack of free will) are all recognized. Recently, attempts have also been

[5] Walter (2001), p. 113.

made to measure free will, for instance by means of neuropsychological tests, especially in ethical and legal contexts.[6]

However, the above theses, implied by the question of whether consciousness is a condition of free will, are inadequate for the issues examined hitherto. It is rather necessary to analyse the consequences of when it is not a condition of free will. Would this mean that consciousness does not play a significant role in the phenomenon of free will? Based on the neuroscientific evidence, the following position is possible. So far, the problem of free will and consciousness has been investigated in relation to human beings or possibly animals. Disregarding some very deep theories rooted in ethics and restricting our analysis to legal purposes, we may say that if someone were not conscious, there would not be any access to her first-person experiences or previous stimuli, or to her mechanisms of making decisions and taking action. This would be so because it is impossible for a human being to see or feel anything that happens in his body, this including the nervous system and the brain, on the cellular or neuronal level. Without consciousness, we could not know whether we ourselves act freely or not, and certainly such information would not be accessible to other people. We could not know whether someone had been at fault if they had acted intentionally or were not in control of their senses. In such a case, the whole concept of law would have been different, as would our world.[7] Hence, from the legal point of view, human consciousness as an element of free will is indispensable; as such, we will treat consciousness as a **condition of free will cognition**.

To complete this introductory picture, it should be also taken into consideration that social institutions have an optimizing function, or rather that in real social life, humans should construct such rules to make their social life possible and improve it. This is indeed the core of all conceptions of the social contract. As a result, overtly utopian legal principles or concepts are sometimes taken for granted as being possible (e.g., the impartiality and objectivity of judges), while concepts which have strong scientific bases (e.g., free will as illusion) must sometimes be treated as false.[8] This is true especially for the law and is the reason why legal concepts are sometimes so different than those of natural sciences or philosophy.

[6] Lavazza and Inglese (2015).

[7] Hyman (2015) and Morse (2015).

[8] This is what Habermas calls an "unavoidable practice". He used this phrase relating to the neutrality principle assessed from the communitarian side: "[. . .] no presumptively neutral principle can ever be neutral in fact [. . .] This objection can be met if one can show that the neutrality principle is a necessary component of a practice that is without alternatives or substitutes, and in this sense unavoidable. A practice is "unavoidable" if it fulfills functions vital to human life and cannot be replaced by any other practice." Habermas (2009), p. 438.

3.2 Free Will and Discernment of AI?

3.2.1 Free Will

As we approach the central issue of this book, it is necessary to consider certain key questions. Is it possible and useful, to acknowledge AI as having free will and discernment? After all, the traditional view on legal subjectivity recognizes them as its indispensable condition. Is such acknowledgement really an indispensable part of legal concepts? Or perhaps free will and discernment should be defined differently in an AI than in a human? These questions will be addressed during the following pages; however, the scope of our interest will be restricted to only weak AI systems, which are not currently conscious, but are becoming increasingly able to solve problems and perform adaptively.

While addressing these questions, it is also worth recalling that according to the new definition of AI given in Article 3 (1) of Proposal 2021:

> 'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

A key part of this analysis is the fragment that an AI system is a software that is able to, *inter alia,* generate predictions, recommendations or decisions, all three of which are inextricably connected with choosing between options: *predictions* select a probable image of a fragment of the world from a pool of images of logically possible worlds, *recommendations* select an image of a fragment of the world from a pool of images of logically possible worlds which allows an assumed objective to be achieved if realized, while *decisions* select an image of a fragment of the world from a pool of images of logically possible worlds which should be brought into existence by a certain action. If the pool used for these three outputs included only one element, neither prediction nor recommendation or decision would have any sense. It is also worth noticing that the techniques and approaches forming part of the definition of AI system, listed in Annex I, which are:

(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
(c) Statistical approaches, Bayesian estimation, search and optimization methods [. . .]

are advanced and flexible enough to gain more complex results than a simple stimulus–reaction relationship.

It must also be considered that a range of AI systems are in use in the contemporary world, from ones which act nearly mechanically (automatically) to more powerful forms based on multilevel artificial neural networks using unsupervised machine learning methods and having access to big data. Furthermore, even the most powerful AI systems do not have to be entirely autonomous, when the burden of

decision making is divided in different parts between AI and its human author, producer, trainer, operator, provider or user. Considering the accelerating pace of development in technology and the fact that today's top achievements may become standard by the time this book is completed, our questions address the kinds of AI which are the most powerful and autonomous of the present time: we cannot foresee the exact directions of development and do not wish to stray into the area of science fiction. Even so, the practical legislative response to our research must account for concrete kinds of AI, with concrete abilities and autonomy. Therefore, the practical legislative solution should comprise a range of contextually situated rules.

Firstly, it is necessary to confirm whether the AI fulfils the **conditions of free will as ability** listed above.

When AI resolves a problem or performs assignments, the number of possible options may be much greater than the number of options given to a human in the same situation. This can be due to many reasons, but an important one is the greater access of AI to big data and its ability to process it: AI systems are typically able to gather more information about the world, and about the relationships between facts, such as causal links, correlations and spurious correlations. Access to such a large data pool gives rise to more options for the AI to achieve its goal. In addition, AI systems are less restricted by the interpretational frames and heuristics which make human thinking rigid; one well-known example is Alpha Go Zero, which passed beyond the limitations of its predecessor Alpha Go by not being constrained by human knowledge:

> The system starts off with neural network that knows nothing about the game of Go. It then plays games against itself, by combining this neural network with a powerful search algorithm. [. . .] After just three days of self-play training, Alpha Go Zero emphatically defeated the previously published version of Alpha Go – which had itself defeated 18-time world champion Lee Sedol – by 100 games to 0. [. . .] Over the course of millions of AlphaGo vs AlphaGo games, the system progressively learned the game of Go from scratch, accumulating thousands of years of human knowledge during a period of just a few days. AlphaGo Zero also discovered new knowledge, developing unconventional strategies and creative new moves that echoed and surpassed the novel techniques it played in the games against Lee Sedol and Ke Jie.[9]

In the case of multilevel artificial neural networks using unsupervised machine learning methods or other equally powerful and autonomous systems, the second condition of **free will as ability**, *viz.* behavioural variability, is by definition fulfilled. The defining goals of artificial neural networks were to learn and solve problems. In the process of learning, an artificial neural network may remove errors in its system by recalibrating the weights of its neuron connections; such recalibration demonstrates that AI can produce different reactions to the same circumstances.

The **conditions of executing free will** are also fulfilled by multilevel artificial networks using unsupervised machine learning methods or equally powerful and autonomous systems. Such systems make decisions autonomously. Except for a given general aim, they are not constrained by rules or strict assumptions which

[9] Silver (2017).

would exclude their making choices between options. They do not even need a model of the world; instead, their model is inferred from sample data, usually by statistical methods. They simply learn by themselves and choose between options of different possible reactions. However, their choices are not random, they are made intentionally to gain the best possible result. Even if the method for choosing the option consists of producing many random decisions and examining which one is the best, it is only a time-saving strategy and cannot be classified as choosing without motive (contrast with the example of Buridan's ass). For example, Amazon's Alexa may choose music or recommend a book according to the taste of the user, answer questions given by the user, or use different languages based on previously-defined user preferences. In turn, systems like Westlaw, a tool of online legal research, offer the following:

> Better set and manage client expectations in terms of cost, timing, and likely outcome by understanding the most probable results. Get a sense of how long it will take to resolve this type of case in front of your judge, how often your judge grants summary judgment motions, and which court is best suited for your client's claim.
>
> Quickly assess and value litigation with Damages, now available in Litigation Analytics on Westlaw Edge. Simply choose your courts and case types from the Damages tab on the Litigation Analytics home page and use appropriate filters to uncover and evaluate awarded damages. Whether you're looking to determine cost-benefit analysis of taking on a case, settlement negotiation approach, or possible client risks or exposures, Damages will help you determine the best course of action.[. . .]
>
> Get the most relevant highlights for your judge, including ruling tendencies, speed, case type experience, appeals, recent activity, and more. Tailor judge analytics data using filters to narrow your results.
>
> Quickly understand the context of your judge compared to the court average, compare judges, or apply new dynamic filters to narrow your results. Plus, apply dynamic filters and control display options.[10]

And finally, **the condition of free will cognition**, which is consciousness. In this point, key differences between human free will and potential AI free will may be observed. For humans, consciousness is the only tool which allows them to learn to make decisions for themselves and others, and the only instrument which lets them report these facts to other people. Without consciousness, humans could not recognize an action and say whether they, or anyone else, acted freely or not. This is quite different to the situation for AI.

Firstly, even unconscious AI may be equipped with software registering or recalling the decision-making process, and which can be used to check whether the choice between options is autonomous. Secondly, in a legal context, it is irrelevant whether AI itself has information about the decision that it made freely. While it is important for reasons of safety for the AI to have access to information regarding whether its decision was free or not (the AI may block unauthorized access

[10] https://legal.thomsonreuters.com/en/products/westlaw/edge/litigation-analytics#expectations, last access on the 4th of August 2022.

to its decisional mechanism), it is more important that humans can access the AI to learn whether its decision was free and without interference: this is needed to verify whether the decision is legally valid, to identify fault or to establish liability. It is humans who are central to human law, and humans should serve as lawmakers, or the judge or may be a victim of fault by an AI. These two elements are together sufficient to confirm that the condition of free will cognition may be fulfilled in an AI. They determine the possible quality of an AI, which may be an excellent substitute for human consciousness for the sake of legal purposes; indeed, the explainability or transparency of AI has been extensively researched, and has long been a postulate.[11]

Many of the documents concerning AI in the legal and ethical domain, for example, GDPR, *High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI*[12] (called in the remainder AIHLEG ETHICS 2019), *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Building Trust in Human-Centric Artificial Intelligence*[13] (called in the remainder "Communication 2019"), Resolution 2020 and Proposal 2021, present formulations of the ethical or legal requirements which should be imposed on AI. A common key requirement in such documents is transparency. The term can be understood as (1) traceability, i.e., logging and documenting both decisions made by the systems and the entire decision-making process, including a description of data gathering and labelling, and a description of the algorithm used; (2) explainability of the algorithmic decision-making process, i.e. explanations of the degree to which an AI system influences and shapes the organizational decision-making process, the design choices of the system and the rationale for deploying it; it also concerns data and system transparency and business model transparency; (3) communication, i.e. the possibility to communicate the capabilities and limitations of the AI system to its users, to identify the AI system thus ensuring that users know they are interacting with one, and to identify the persons responsible for it (Communication 2019, p. 5). In legal or prelegal documents, *transparency* is often referred to as *explainability*. As a consequence, one can differentiate between explainability *sensu largo*, i.e., as a synonym of transparency, and explainability *sensu stricto*, as in point (2) above. The reminder of this book will assume the former meaning of the term. It is however important not to confuse either notion with that used in computer sciences, where they speak of executive traces, explanations, interpretations and justifications. *Traces* identify which statements are being

[11] About the explainability and transparency the authors wrote in the context of causality: Wojtczak and Księżak (2021). The fragment below is taken from this article.

[12] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, 8 April 2019. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419, last access on the 4th of August 2022.

[13] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Building Trust in Human-Centric Artificial Intelligence, Brussels, 8.4.2019, COM (2019) 168 final.

executed in the operation of the program. *Interpretations* are descriptions of these operations, which can be understood by a human, and can be divided into *explanations* and *justifications*. *Explanations* report how a given decision was made and show causations and correlations, while *justifications* explain why decisions are correct. [14]

In addition to the legal regulations concerning AI, some requirements have been formulated in Computer Science and by businesses; however, they may not be created for ethical or legal purposes. The creation of such guidelines is often justified by the belief that they build safety and trust, and support the development of computer technologies. By following such guidelines, AI can be effectively applied to serve various marketing purposes, detect fraud, illegitimate transactions or identity theft, and can predict and identify anomalies in various domains.[15] Hence, a new discipline in Computer Science was established, *Explainable Machine Learning Approaches, Explainable Artificial Intelligence* or *XAI*, whose purpose is to clarify AI decision-making, its actions and its recommendations. Based on its findings, it should be possible to move from confusion and concern ("Why did you do that? Why not something else? When do you succeed? When do you fail? When can I trust you? How do I correct an error?") to certainty and confidence ("I understand why; I understand why not; I know when you succeed; I know when you fail; I know when to trust you; I know why you erred").[16]

However, although such ethical, legal and utilitarian requirements seem right and reasonable, they are difficult to follow with existing technology, and it is likely that achieving compliance will become significantly more difficult in the future. According to one of the reports prepared by the EU regarding the law of robotics:

> AI methods are famously known to have limited capacity to provide the reasoning principles behind a decision, mainly due to the fact that the logic is automatically inferred from vast amounts of data, and embedded in complex mathematical structures that are successful but very opaque for humans. The explainability of methods is then becoming crucial in this context to ensure the rights of individuals to understand decisions concerning them.[17]

Furthermore, Blanco-Justicia and Domingo-Ferrer (2019), p. 15 clearly state that:

> [. . .] there is a risk of automated decisions becoming an omnipresent black box. This could result in formally transparent democracies operating in practice as computerized totalitarian societies.

Despite many attempts to protect the legal or ethical rights of citizens to obtain the explanations for the decisions that affect them, today's AI systems have become too complicated to be understood by humans. A complex analysis of the legal and technical feasibility of the explainability of algorithmic decisions by Brkan and Bonnet (2020), pp. IV.2b–IV.2c found that while such explainability is not generally

[14] Brkan and Bonnet (2020), pp. II.2–II.3.

[15] simMachines, https://simmachines.com/explainable-ai/, last access on the 4th of August 2022.

[16] Turek (2016).

[17] Hamon et al. (2020).

impossible from the perspective of Computer Science, particularly XAI, success is very difficult to achieve in practice; in this case an "impossibility result is a formal proof showing that a given problem cannot be solved by a given computational model". A more thorough explanation was given by Blanco-Justicia and Domingo-Ferrer (2019), p. 16:

> To be scalable, explanations must be automatically generated: even if a human auditor was able to produce a compelling explanation, one cannot assume that such an auditor will be available to explain every automated decision to the affected subject. Older machine learning models, based on rules, decision trees or linear models, are understandable by humans and are thus self-explanatory, as long as they are not very large [. . .] However, the appearance of deep learning has worsened matters: it is much easier to program an artificial neural network and train it than to understand why it yields a certain output for a certain input.

Furthermore, several risks are associated with generating explanations which are accurate, consistent, stable, representative, certain, novel and guarantee fidelity. The first is the risk of revealing such information about the training data set or the system itself, which may be a trade secret. This entails a further risk of revealing classified information, ranging from top secret to restricted, associated with patents and copyrights: "a software and a source code can be copyrighted, but not an algorithm that is merely an abstract idea underpinning the software and the source code".[18] In addition, such explanations run the risk of revealing information containing personally-identifiable elements encoded in the training data set. A further risk is associated with the comprehensibility of the data, i.e., the explanation itself may not be comprehensible to humans;[19] importantly, such an "explanation is different for someone who is end-user, a developer, or an external affected or forced to interact with an autonomous system".[20] This is why IEEE[21] P7001 standard[22] defines five distinct groups of subjects for which AI systems must be transparent in different ways and for different reasons: end users, wider public and bystanders, safety certifiers, incident/accident investigators, lawyers and expert witnesses.[23] Furthermore, the nature of the explanation depends to a great degree on the person generating it: the owner of the AI system and training data set may limit access to the system to protect trade secrets, and may even "induce" opaqueness of the system,[24] while the third party or the recipient himself may be more transparent.[25] Hence, the explanation is, and should be, drawn up for the context in which it is needed, i.e., there cannot be a single universal explanation added to every AI

[18] Brkan and Bonnet (2020).

[19] Blanco-Justicia and Domingo-Ferrer (2019), pp. 17–18.

[20] Glomsrud et al. (2019).

[21] Institute for Electrical and Electronics Engineers (IEEE) https://www.ieee.org/about/at-a-glance.html, last access on the 4th of August 2022.

[22] IEEE (2020), pp. 1–76.

[23] Winfield et al. (2021).

[24] Monterossi (2019), p. 717.

[25] Blanco-Justicia and Domingo-Ferrer (2019), p. 25.

product, as is the case with a user manual; unfortunately, this fact may be a source of many difficulties.

These concisely-described limitations of transparency may represent just some of the potential sources of legal problems, for instance when attempting to identify a causal link or attributing guilt or action. As mentioned earlier, such problems are not new and did not appear together with the development of AI: early examples can be seen in the transition from preindustrial to industrial societies. They have also been associated with the growing complexity of industrial processes:

> [. . .] the amplifying effect of the complexity was not resolved in a mere multiplication of the number of accidents involving damage. The transformation also regarded their intrinsic quality. Such new damaging facts more and more often were connected to technical and industrial data: the progressive consolidation of interaction between humans and machines in the process of industrial production made it hard to define the source from which the damaging facts emerged. Their matrix, in other words, become anonymous and the causal connection between a specific action and its outcomes more difficult to be identified and proved.[26]

However, the "analogous world" is easier for humans to perceive and understand than the digital one; this is the case for many reasons, but most importantly, that the majority of processed data and machine code used for processing by AI is strictly numerical. As such, the contemporary problems generated by limited transparency are both qualitatively and quantitatively more significant.

There is also another very specific problem with the explainability of AI systems: improving such explainability necessitates further programming of the AI system, resulting in additional costs, and forcing greater openness of the AI system may challenge the producer's monopoly on the product, resulting in lower gross income. Indeed, some organizations, such as European Digital Rights (EDRi) in the Consultation Response prepared to White Paper 2020, even propose that:

> There should be liability for producers of AI that do not disclose source code (including their algorithmic models / datasets) and do not provide fixes for issues brought to their attention or otherwise hinder fixes from being applied, for example by not allowing third-party fixes based on any disclosed source code.[27]

Sometimes also, as concisely put by Laber and Murtinho (2021), "the price of explainability can be thought of as the loss in the terms of quality that is unavoidable if we restrict these systems to use explainable methods".

As such, the demand for explainability or transparency is, especially from the business point of view, rather unpalatable. This may be the reason why recently we have seen the drive for explainability or transparency to ease a little. Proposal 2021 (p. 31), indicates:

[26] Monterossi (2020), p. 5.

[27] EDRi (2020), European Commission Consultation on the White Paper on Artificial Intelligence. EDRi Consultation Response, p. 7. https://edri.org/wp-content/uploads/2020/06/AI_EDRiConsultationResponse.pdf, last access on the 4th of August 2022.

> (47) To address the opacity that may make certain AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for high-risk AI systems. Users should be able to interpret the system output and use it appropriately. High-risk AI systems should therefore be accompanied by relevant documentation and instructions of use and include concise and clear information, including in relation to possible risks to fundamental rights and discrimination, where appropriate.

However, in this part of the document, and some others, e.g., Resolution 2020, all the significant demands focus only on so-called high-risk AI systems, and transparency obligations are boiled down to informational obligations for exposed natural persons (Article 52 Proposal 2021). This self-restraint by the European Commission is justified by the principle of proportionality as well as the necessity to obtain the objectives of the regulation, which

> follows a risk-based approach and imposes regulatory burdens only, when AI system is likely to pose high risk to fundamental rights and safety. For other, non high-risk AI systems, only very limited transparency obligations are imposed [. . .]
>
> (Proposal 2021, part 2.3 of Explanatory Memorandum, p. 7)

However, if the hopes of the legislative bodies and various nonpublic institutions regarding greater explainability are realized, this success may be used for legal purposes in constructing a specific notion of free will in AI, and the problem of lack of consciousness would be solved. However, this does raise the question of what this notion could look like, and how it can be used. There would also be a need to think of a new name for the equivalent of free will. Although names are not the most important considerations, "free will" is so philosophically meaningful that it may be useful to get rid of it when we speak of AI. Among many possible names, we propose the title "unhindered competence of deciding" (in the remainder UCD).

The AI system has UCD (substitute of human free will) if altogether:

1. AI and its characteristics, among others, its aim ("intended purpose" according to Proposal 2021) and mechanism of reasoning, are legally registered/certified;
2. AI and its actions are fully explainable;
3. it learns in an unsupervised way, or the process of supervised learning is finished;
4. no human or another AI may influence the process of data gathering or the mechanism of choosing;
5. the data necessary to make the best possible choice are gathered while unsupervised, and the AI has unsupervised access to the best possible sources (this condition includes the hardware requirements necessary for gathering data, e.g. sensors, processors, meters etc.); the possibility is evaluated according to the given circumstances;
6. it uses adequate mechanisms of data verification and evaluation;
7. it finds at least two options;
8. it chooses one of the options, being motivated by the registered aim;
9. it acts according to the made choice;
10. its choices and actions are justified by its aim, gathered data and registered mechanism of reasoning.

3.2.2 Discernment

Linking free will and discernment is fully justified when one remembers that the choice to be free must be motivated and not random or even, as some authors say, rational. We speak about it above, calling this demand a **condition of executing free will.**

Thus, AI, to execute free will or to be a fully-fledged subject of civil law, should also be endowed with discernment. Admittedly, the issue of discernment is noticed when the capabilities of human beings are at stake, and not, for instance, of corporations; this is so because there are certain legal preliminary conditions regarding the enacting and functioning of juridical persons, which assume that a corporation has discernment throughout the whole of its existence, i.e. from its beginning to its end. For example, the board of directors in a corporation must consist of fully legally-capable persons: children cannot have a seat. Furthermore, within tort law and fault liability, it is a principle in Western legal systems that the juridical person is obliged to redress the damage caused by its bodies and, usually, for which these bodies were at fault. This is because the members of bodies are humans and they have their own free will and discernment which collectively (although not in total) are the substitute of the will and discernment of a juridical person.

Discernment influences both the duties (obligations) and the rights (competences) of humans. It is usually standardized to some degree by the law, which associates it with the age of a natural person and his or her full or partial incapacitation, having its source in physical or mental health. Discernment may be defined and named in different ways. We think the approximate equivalent of discernment may be reasonableness. It is true that this notion usually serves its purposes, particularly as it gives the standard of expected behaviour accepted in civil law, but it should be noted that this standard can only be fulfilled by the person to whom discernment may be attributed. DCFR, for instance, recommends that this standard should be objective but contextually situated:

> **1.– 1:104 Reasonableness**
> Reasonableness is to be objectively ascertained, having regard to the nature and purpose of what is being done, to the circumstances of the case and to any relevant usages and practices.

From the perspective of the problems given above, this definition could imply that a person who is endowed with discernment should be capable of discerning what is reasonable, from an objective point of view, with regard to the nature and purpose of what is being done, as well as the circumstances of the case and any relevant usages and practices.

Therefore, in this respect, could AI be regarded as being more like a juridical person or an individual? For the juridical person discernment is assumed a priori when the juridical person meets the conditions of the given legal system, for human it depends on his individual characteristics (e.g. age, sanity etc.) And is the

discernment an element which should be taken in consideration at all? We believe that it depends on the part of the civil law which is to be applied. If AI acts on the field of contract law, the question of discernment does not arise at all, because when the law permits such an action, i.e., it endows AI with some legal subjectivity, the law determines also the preliminary conditions which sift a "discerning" AI, which is allowed to act, from a "not-discerning" one, which is not. These preliminary conditions may be certification and registration. In this case, AI is more like a juridical person. However, in the case of tort law, AI is more like a human, because, despite being a system composed of many parts, these parts belong together, and no matter whether AI is embodied in some mechanical form or not, its operation is similar to that of an organism. If AI causes damage, the tort-feasor is the AI itself and not its parts, bodies or auxiliaries.

Of course, it does not mean that the law needs to specify an age limit or conditions of incapacitation. These are substituted for by the preliminary conditions of participation in legal facts. However, some problems may arise, and these resemble those concerning human discernment.

References

Books and Articles

Blanco-Justicia A, Domingo-Ferrer J (2019) Machine learning explain ability through comprehensible decision trees. In: Holzinger A, Kieseberg P, Min Tjoa A, Weippl E (eds) Machine learning and knowledge extraction. Springer, Cham. ISBN 978-3-030-29726-8

Brkan M, Bonnet G (2020) Legal and technical feasibility of the GDPR's quest for explanation of algorithmic decisions: of black boxes, white boxes and fata morganas. Eur J Risk Regul. 11(18): II.2–II.3. ISSN 2190-8249

Glomsrud JA, Ødegårdstuen A, St. Clair AL, Smogeli Ø (2019) Trustworthy versus Explainable AI in Autonomous Vessels. Conference: ISSAV 2019 - International Seminar on Safety and Security of Autonomous Vessels At: Hanasaarenranta, Espoo, Finland: 2019, https://www.researchgate.net/publication/336210763_Trustworthy_versus_Explainable_AI_in_Autonomous_Vessels, last access on the 4th of August 2022

Habermas J (2009) Between facts and norms. Polity Press, Cambridge

Hamon R, Junklewitz H, Sanchez I (2020) Research Centre Technical Report. Robustness and explainability of artificial intelligence – from technical to policy solutions. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/57493. (online), JRC119336

Hart HL (1994) The concept of law. Oxford University Press, New York

Hill TT (2019) Neurocognitive free will. Proc R Soc B 286:20190510. https://doi.org/10.1098/rspb.2019.0510

Hyman SE (2015) Neurobiology collides with moral and criminal responsibility: the result is double vision. In: Glannon W (ed) Free will and the brain: neuroscientific, philosophical, and legal perspectives. Cambridge University Press, Cambridge

Kane R (1998) The significance of free will. Oxford University Press, New York

Laber ES, Murtinho L (2021) On the price of explainability for some clustering problems, preprint, https://www.researchgate.net/publication/348251411_On_the_price_of_explainability_for_some_clustering_problems, last access on the 4th of August 2022

Lavazza A (2016) Free will and neuroscience: from explaining freedom away to new ways of operationalizing and measuring it. Front Human Neurosci 10:Article 262. https://doi.org/10.3389/fnhum.2016.00262

Lavazza A, Inglese S (2015) Operationalizing and Measuring (a Kind of) Free Will (and Responsibility). Towards a new framework for psychology, ethics, and law. Rivista Internazionale di Filosofia e Psicologia 6(1):37–55. https://doi.org/10.4453/rifp.2015.0004

Libet B (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. Behav Brain Stud 8(4):529–566

Monterossi MW (2019) Algorithmic decisions and transparency: designing remedies in view of the principle of accountability. Ital Law J 5(2) ISSN 2421-2156

Monterossi MW (2020) Liability for the fact of autonomous artificial intelligence agents. Things, agencies and legal actors Global Jurist 20190054, eISSN 1934-2640, https://doi.org/10.1515/gj-2019-0054

Morse SJ (2015) Neuroscience, free will, and criminal responsibility. In: Glannon W (ed) Free will and the brain: neuroscientific, philosophical, and legal perspectives. Cambridge University Press, Cambridge

Silver D, Schrittwieser J, Simonyan K et al (2017) Mastering the game of Go without human knowledge. Nature 550:354–359. https://doi.org/10.1038/nature24270

Turek M (2016) Explainable Artificial Intelligence (XAI). Defense Advanced Research Projects Agency (DARPA). United States Government. https://www.researchgate.net/deref/https%3A%2F%2Fwww.darpa.mil%2Fprogram%2Fexplainable-artificialintelligence%23%3A%257e%3Atext%3DXAI%2520is%2520one%2520of%2520a%2Cto%2520characterize%2520real%2520world%2520phenomena. Last access on the 21st of October 2022

Walter H (2001) Neurophilosophy of Free Will: from libertarian illusions to a concept of natural anatomy. The Massachusetts Institute of Technology

Winfield A, Booth S, Dennis LA, Egawa T, Hastie H, Jacobs N, Muttram RI, Olszewska JI, Rajabiyazdi F, Theodorou A, Underwood MA, Wortham RH, Watson E (2021) IEEE P7001: A Proposed Standard on Transparency. Front Robotics AI 8:665729. https://doi.org/10.3389/frobt.2021.665729

Wojtczak S, Księżak P (2021) Causation in civil law and the problems of transparency in AI. Eur Rev Priv Law 29(4):561–582

Documents

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Building Trust in Human-Centric Artificial Intelligence COM (2019)168 final. 8.04.2019. Brussels. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2019:168:FIN, last access on the 4th of August 2022

European Commission Consultation on the White Paper on Artificial Intelligence: EDRi Consultation Response. 4 June 2020. Brussels. https://edri.org/wp-content/uploads/2020/06/AI_EDRiConsultationResponse.pdf, last access on the 4th of August 2022

IEEE Draft Standard for Transparency of Autonomous Systems, in IEEE P7001/D1, June 2020. (Piscataway, NJ: IEEE), 1–76. https://www.techstreet.com/ieee/searches/35199403, last access on the 4th of August 2022

Online Sources

simmachines. https://simmachines.com/explainable-ai/, last access on the 4th of August 2022

Chapter 4
Capacity for Juridical Acts

4.1 Introduction

This chapter will examine the concept of capacity for juridical acts and discuss it in the context of the participation of AI in legal relations. It should be noted, however, that different countries and their legal systems use different notions under different names and with different scopes; these notions are connected to the possibility of participation in legal relations, and the capacity for juridical acts (or perhaps **active juridical capacity**) is only one of them. For instance, in some legal systems, a specific term is used to denote the capacity to be a party to legal relations in general, and to be attributed with legal obligations and legal rights. Such a capacity is usually recognized as a preliminary condition of capacity for juridical act. The term used for this concept can be, for example, *juristic capacity* or *legal capacity*; however, because these terms are synonymous with *legal personality* or *legal subjectivity*, and in fact these notions are analysed in Chap. 2, they will not be touched here, except for when such references are needed to realize the objectives of this chapter. This chapter is devoted the capacity for juridical acts.

A contract, or any other juridical act, can be valid only when it is concluded by a subject of the law who has the legal capacity for this juridical act. Among humans, usually two categories of people qualify as not being capable for juridical acts at all: certain minors, young children in particular, and people with mental disorders, who are fully incapacitated. The reason for such regulation is that such persons, although they are humans and legal subjects, have not enough discernment to care about their own interests and cannot properly use and express their will, which should be free will. The problems of discernment and free will are generally discussed in Chap. 3. This chapter will elaborate more particular issues, especially the consequences of accepting the concepts of certain discernment and free will.

When a natural person is excluded from conscious and free decision-making, and from the conscious and free expression of his will, due to his state, the juridical acts

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_4

of this person are recognized by the law as defective. This defectiveness may be defined in different ways and its effects may be different in each legal system.[1] In civil law countries, any contracts concluded by a person without sufficient mental capacity are usually void *ex lege*, and in common law countries, they may be invalidated. These differences are not needed in our further analysis, nor are the structural differences between different regulations of the capacity for juridical acts. As noted previously, the capacity for juridical acts is generally not applicable to those under a certain age, or those who have been deprived of such by a judicial decision based on mental disability and the need for help in arranging aspects of life. In both cases, the aim of this imperfect legal status is not to punish or limiting someone's rights, but to protect a person who, acting by his own, could become a victim of various detriments. These rules are typically applied for a given time, but only until the need for protection and help ceases, for example, when a person reaches the age of majority or when he or she recovers, and the court cancels the state of incapacitation. Such a model of treatment of persons with disabilities was disseminated and empowered by Recommendation No. R (99) 4 of the Committee of Ministers of the Council of Europe to Member States on Principles Concerning the Legal Protection of Incapable Adults (1999) and Convention on the Rights of Persons with Disabilities and Its Optional Protocol,[2] especially Article 12, which requires the states—parties of this agreement to:

- recognize persons with disabilities as persons before the law, enjoying equal legal capacity as others in all aspects of life,
- taking appropriate measures to provide access by persons with disabilities to any support they may require in exercising their capacity,
- ensure that all measures that relate to the exercise of legal capacity provide for appropriate and effective safeguards to prevent abuse in accordance with international human rights law. Such safeguards shall ensure that measures relating to the exercise of legal capacity respect the rights, will and preferences of the person, are free of conflict of interest and undue influence, are proportional and tailored to the person's circumstances, apply for the shortest time possible and are subject to regular review by a competent, independent, and impartial authority or judicial body. The safeguards shall be proportional to the degree to which such measures affect the person's rights and interests.
- take all appropriate and effective measures to ensure the equal right of persons with disabilities to own or inherit property, to control their own financial affairs and to have equal access to bank loans, mortgages, and other forms of financial credit, and shall ensure that persons with disabilities are not arbitrarily deprived of their property.

[1] Smits (2014), pp. 91–100.

[2] Convention on the Rights of Persons with Disabilities and Its Optional Protocol (A/RES/61/106) adopted on 13 December 2006 in New York, entered into force on 3 May 2008, https://www.un.org/disabilities/documents/convention/convention_accessible_pdf.pdf, last access on the 4th of August 2022.

This model is known as the *social model*, which assumes that "people are viewed as being disabled by society rather than by their bodies", as opposed to the former *medical model*.[3] The change of models also implies moving from substituted decision-making towards supported decision-making. [4]

Depending on the age and the kind of incapacitation (e.g. partial or full, as observed in Slovakia, Slovenia and Poland, as well as other kinds), the way in which a person lacking the capacity for juridical acts may become a party to a contract or make a declaration of intent may differ; for example, this may occur through the action of a statutory representative or curator appointed by the court—the names and models of guardianship differ, or with the permission of the statutory representative or curator appointed by the court. In addition, the scope of juridical acts which a person may make on his own may vary. In Poland, for instance, they may be allowed to conclude contracts within petty, current matters of everyday life, while in Austria these can be everyday transactions of little significance, and in Estonia, transactions from which no direct civil obligations arise for the person.[5]

However, a separate legal institution is usually provided for cases when a person who has a legal capacity for juridical acts, i.e., one of proper age and not incapacitated, may lose the ability to make decisions freely or consciously for some time, even momentary. This may happen because of consuming alcohol or drugs, or because of some momentary disease or disorder symptom. Such a state is usually described in detail in the local legal rules. In such cases, any juridical act performed by a person being in such a state at the moment of its concluding, is not valid. For example, in the Polish Civil Code, it is written as follows:

> Article 82. A declaration of intent shall be invalid if it was made by a person who for whatever reason was in a state excluding conscious or free decision-making and expressing his intent. It shall in particular concern a mental illness, mental retardation or other, even a temporary, mental disorder.

While in BGB:

> § 104. A person is incapable of contracting if
>
> 1. he is not yet seven years old,
> 2. he is in a state of pathological mental disturbance, which prevents the free exercise of will, unless the state by its nature is a temporary one.

Certainly, the question arises how the second party may recognize a state of this kind, and how to prove that a person was in such a state at the moment of giving a declaration of intent. However, this is an issue of adequate measures of inquiry.

---

[3] World Health Organization (2011) World Report on Disability, Geneva, WHO, p. 4. after FRA, Legal capacity of persons with intellectual disabilities and persons with mental health problems, Vienna – Austria 2013.

[4] FRA, Legal capacity of persons with intellectual disabilities and persons with mental health problems, Vienna – Austria 2013, p. 27.

[5] Varul et al. (2004).

What was said above relates to the capacity for juridical acts of natural persons. The same capacity for juridical persons (sometimes called *competence* or *power*) seems to be less complicated at the general level. Juridical persons are capable of juridical acts within the scope determined by the legislation, the deed of incorporation, articles of association or other constitutional documents and implied by the nature of juridical persons in general. Not all juridical acts which are possible for human beings are included in this set. For instance, juridical persons, because they are not biologically living and cannot die, cannot make a last will. However, except for such particular cases, juridical persons may perform all juridical acts, unless special circumstances arise which make performing the juridical act impossible (e.g., temporary lack of bodies).

4.2 Capacity for Juridical Acts of AI: Theoretical and Legal Bases

To build some theoretical bases for further discussion, firstly, it is necessary to review the classification of facts. On the first level, they can be divided into non-juridical facts (sometimes called *natural*, although this term is not precise enough) and juridical facts. Such natural facts are concrete states of affairs which actually happen in the "ordinary" world and do not influence legal relationships. Juridical facts are concrete states of affairs which trigger a reaction of the law, which cause some legal consequences, which count as changes in the world of the law.[6] We will not describe the classification of these natural facts in more detail because both classifications of non-juridical facts and juridical facts run almost parallel to each other.

On the second level, juridical facts may be classified as juridical occurrences and juridical conducts. Juridical occurrences are juridical facts which happen independently of human will, such as birth, death and unjust enrichment; in contrast, juridical conducts are juridical facts which happen as a result of human will. Juridical conducts may be further divided into juridical actions and juridical acts. Juridical actions are examples of human conduct which, although dependent on human will, are not made with the intention of inducing any legal consequences. They may be legal or illegal. The best examples of illegal juridical actions are torts, while those of legal juridical actions may be taking in a dog or finding another person's possessions. In contrast, juridical acts are juridical conducts undertaken with the intent of causing certain legal effects. According to the definition of Hage (2011b), pp. 49–50

> Juridical acts are acts to which the law connects legal consequences. The characteristic that sets of juridical acts from other acts with legal consequences is that the consequences of a typical juridical act are those which actor wanted to bring by the means of his act. [. . .] a

[6] Hage (2011a), pp. 33–34.

> juridical act is an act performed with the intention to create legal consequences and to which the law attaches the intended legal consequences because they were intended.

According to the definition of the DCFR:

> II. – 1:101(2) A "juridical act" is any statement or agreement, whether express or implied from conduct, which is intended to have legal effects as such. It may be unilateral, bilateral or multilateral.

The DCFR also gives the definition of a contract which is recognized as the most typical example of juridical act:

> II. – 1:101: (1) A contract is an agreement which is intended to give rise to a binding relationship or to have some other legal effect. It is bilateral or multilateral juridical act.

This well-known classification clearly shows that while considering the potential capacity of an AI for juridical acts, it is essential to settle whether and how an AI may have something which may be called "will" (will is necessary for intent) and whether it may express the will in some legally accepted way. Indeed, it is not possible to conclude a juridical act without the concrete intent of actuating certain precisely-determined legal consequences; in this sense, *a concrete* means to be present at a certain time and in a certain place by a certain entity and of certain content. Concrete intent can only be established based on the general capacity of will and some concrete knowledge about the world, such as the operation of a given legal system, as a result of the general capacity of discernment.

Our current level of technological development does not give any reason to think that AI (weak AI) could be conscious in the near future; as such, AI cannot be regarded as being equally capable of juridical acts as natural persons. We cannot attribute discernment and free will to an AI in the same way they are recognized in human beings. This issue was debated in this monograph in Chap. 3. Of course, it is not physically or logically impossible that this situation can change over time. The development of strong (wide, general) AI may well hasten the technological advancement needed for it to gain consciousness. In the legal domain, this would result in a complete U-turn and the need for the discussion to begin anew. Despite this, AI cannot be regarded as equivalent to a less skillful than an average human being, such as a minor or a mentally-disabled adult. As it was shown above, such humans, because of their dignity, are not deprived of their right to participate in legal transactions but are limited only to the extent which is necessary to help and protect them. Of course, there is no reason to hold such an ethically careful attitude towards AI.

However, even today the notion of capacity for juridical acts related to AI is not entire nonsense. As it was demonstrated in Chap. 3, the autonomy of AI does not allow it to be regarded as an ordinary human tool. While the tool serves only as a means to transfer human will, or metaphorically as an extension of a human arm (whether it is a ballpen or an advanced algorithm), there is no sense, or any need, to consider whether the tool has any cognitive capabilities; furthermore, it is pointless to discuss whether it has any point of view regarding the motivation of its user. However, this state of affairs changes dramatically with the presence of advanced

AI. Its autonomy, even if unconscious, in the external world gives results which are truly based on decision making. The question is whether the law should recognize these decisions as legally binding, which would mean that the law would recognize the capacity of AI for juridical acts.

As it was explained in Chap. 3, AI satisfies all conditions of having free will except for one, which is consciousness. However, we assert that from the perspective of the law, consciousness is important only for natural persons because natural persons cannot control and report the process of making decisions in any other way; furthermore, they cannot without consciousness say whether this process was unhindered or not. Without consciousness, people are "black boxes"; they are entirely unexplainable.

The significance of this element is especially confirmed by the fact that the law demands also that the decisions of juridical persons be explainable; hence minutes are taken of board meetings or associate assemblies, with the presence of a notary public being sometimes legally required on these assemblies, and for reports to be kept in statal registers.

Admittedly, we doubt that it is possible to provide the full explainability/transparency of AI. Nevertheless, we cannot ignore that it is postulated and assumed by different official and *ad hoc* bodies participating in the discourse on AI legal status. Most believe that explainability/transparency may be obtained either by giving AI access to its internal processes and programming it to report its own decision-making course, or by giving the task of overseeing AI decision-making processes to some external entity, most often human but also another AI. There are also proposals to regulate this issue explicitly by legal rules. However, some regard the demands of explainability as being too costly for business, and would detract from the expected development of AI.[7]

So, we say conditionally that if the proper degree of explainability of AI is attained, and it can be ascertained that AI's decisions are unhindered (AI has UCD[8]), a lack of consciousness should not be an obstacle to endowing AI with capability for juridical acts.

Besides, if the explainability of a given AI system were to entail the possibility of it obtaining capacity for juridical acts, this may be a nudge for AI producers to develop explainable AI systems, at least for certain purposes. And it would be a very good, although not expected side-effect of the legal regulation within the domain of civil law.

The second element needed to obtain capacity for juridical acts is discernment. As noted in Chap. 3, this is not a problem for contract law at all. If AI is effective and competent enough to perform certain actions without the supervision of the user, this fact will be confirmed by the producer, by the market and by the statal certificate or register record. This state is analogous to that of juridical persons. A juridical person must fulfill the necessary legal requirements: for instance, all its bodies are appointed

[7] Cf. Chap. 3.

[8] Cf. Sect. 3.2.1 in fine.

in the legally-determined composition, it is registered, and automatically has the capacity for juridical acts. In this way, the law recognizes that a juridical person has sufficient discernment to act.[9]

However, here we should observe a certain difference: unlike a juristic person,[10] but similarly to a natural person, an AI system may be autonomous either within the scope of creating natural facts or, if the legal system gives such a capacity to AI, within the scope of juridical acts. If an autonomous vehicle delivers a passenger to the requested place by adjusting the route to account for some parameters (assuming that the vehicle is sufficiently autonomous to select one of two equally good routes rather than stopping), its autonomy demands a capacity to create natural facts which are moving from point A to point B. Alternatively, if an autonomous trading system closes an auction sale and, based on capacity for juridical acts given by the legal system, awards some participant with the rights of an emptor, it creates or concludes a juridical fact. Such a situation would not be anything exceptional and technically is quite possible, indeed various automated (mechanical) trading systems are quite popular today.[11] This kind of system is regulated by the MIFID II.[12] According to Article 4.1 of this act:

> (39) 'algorithmic trading' means trading in financial instruments where a computer algorithm automatically determines individual parameters of orders such as whether to initiate the order, the timing, price or quantity of the order or how to manage the order after its submission, with limited or no human intervention, and does not include any system that is only used for the purpose of routing orders to one or more trading venues or for the processing of orders involving no determination of any trading parameters or for the confirmation of orders or the post-trade processing of executed transactions;
>
> (40) 'high-frequency algorithmic trading technique' means an algorithmic trading technique characterised by:

[9] Within tort law, the problem is more complicated: having the capacity for juridical acts is less important than the capacity to be personally responsible for damage. However, these two capacities are to a certain extent related. This relation is implied by the fact that discernment and free will may be acknowledged as a necessary element of fault or responsibility. In the case of natural persons, the law may also accept chronological age/bright-line test with certain thresholds of age, or a subjective test based on the capacity of a particular child to recognize and avoid risk and harm; this test takes into account age, intelligence and experience.

[10] AI embedded in a robot may move physical objects, move itself, generally use the physical power. Juristic persons cannot use the physical power by themselves, they must use humans or tools for it.

[11] Trading systems are often divided into mechanical and discretionary types. The names may be confusing, because "mechanical" means that almost all, if not all, trading decisions are delegated to the system, while a discretionary system assumes human participation. Mechanical trading systems are commonly referred to as algorithmic trading systems or automated trading systems. Hasan (2021), Wood and Sutphen (2015).

[12] Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (recast) L 173/349. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0065, last access on the 4th of August 2022.

(a) infrastructure intended to minimise network and other types of latencies, including at least one of the following facilities for algorithmic order entry: co-location, proximity hosting or high-speed direct electronic access;
(b) system-determination of order initiation, generation, routing or execution without human intervention for individual trades or orders; and
(c) high message intraday rates which constitute orders, quotes or cancellations;

The use of these kinds of trading systems, which are considered very risky (especially high-frequency algorithmic trading), is regulated in Article 17 of MIFID II; this article imposes various obligations on an investment firm engaging in algorithmic trading to ensure, speaking generally, the safety of markets and clients. However, this document does not appear to include any analysis of the legal status of the AI systems which are the real makers of the concluded transactions.

Proposals have been made to explain their action in legal terms, but they are not entirely satisfactory. The initial theories on the role of AI systems in the course of electronic transactions arose as a result of the development of e-commerce. The issues important for electronic contracts are described in Chap. 9. Here it is sufficient to give Habibizadeh's (2016, p. 3) list of possible theories.

The main theories are as follows:

(a) The electronic agent is merely a tool of communication between the parties and nothing more.
(b) Unilateral offer theory, this does not pay attention to the status of electronic agents but rather it analyses the nature of display of goods and services on a website.
(c) Ignorance of the satisfaction of a particular intention for each contract without any need for intention.
(d) Applying the objective test to justify the existence of a legal intention in contracting electronically.
(e) An electronic agent is considered to have a legal personality.
(f) An electronic agent is the agent of the user under the law of agency.
(g) An electronic agent is the agent of the user without a legal personality: theory of slavery.

Although the above list seems to be a long one, it can be boiled down to three competitive positions.

Firstly, it is possible to insist that AI has no legal capacity or capacity for juristic acts, because computer software cannot be classified in such categories. Such characteristics are not adequate nor useful for describing the legal effects of using AI. This opinion could be associated with theories a) b) c) and d) on Habibizadeh's list. However, it can be refuted because of the reasons given above. If it is assumed that the contracts concluded by electronic means are valid, and this assumption is legally and practically indefeasible (Article 9. 1 of Directive on electronic commerce[13]), it is necessary to determine the legal role of such means in concluding these contracts. If electronic means are used to choose whether the contract is

[13] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), L 178/1, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031, last access on the 4th of August 2022.

concluded or not, and identify the contractor and its essential elements, the criteria regarding when such choices bring legal effects must be delineated. Of course, this demand does not automatically mean that the criteria must form any human-like capacity: they may resemble that awarded to a juridical person or may even be an entirely different new concept.

Secondly, it is possible to consider that any AI which acts in legal transactions should always have legal capacity and capacity for juristic acts, if these attributes are necessary to validate the juridical act which the AI must conclude. This idea would be analogous to that of funding the participation of juristic persons in the legal order. Juristic persons, although not actually capable of performing many usual human actions, are never treated as minors or legally incapacitated adults. Their legal capacity is treated as defaulted; it is unquestioned and the scope of their capacity for juristic acts is implied by legal rules. Indeed, it may sometimes be useful to search for similarities between AI and juristic persons, as noted in Chap. 2. However, this analogy does not work in cases where it is important that humans are acting for a juristic person, and it is their will which counts individually or collectively. When, for example, the president of the board concludes a juristic action while mentally incapable, or is misinformed, the validity of this action could be questioned on the grounds of vitiated consent or intention. If AI were a contractor, the situation would be different, because it would be AI which made the autonomous decision, and only the circumstances relating to this AI could be used to justify the invalidation of a juristic act.

In addition, each legal system establishes rules which allow given organizational units to be classified as legal persons; usually the type of organizational unit is precisely described and it is necessary to include this type in a certain register under a specific business name. However, in the case of AI, it is difficult to discuss the problem of its capability for juristic acts because there are no indispensable legal instruments to differentiate between forms of AI which could be capable of participating in legal transactions and other forms of AI or software. There is no universally-accepted definition of AI, there is no proper register with proper records, and there is no standard nomenclature for AI which can attribute it a unique, easily comprehensive symbol. While a significant and official proposal exists for such a definition and register (Proposal 2021), this may change during consultations and the legislative process and it is not clear whether the intention of its creators was to use the proposed rules within civil law. It is worth noting again at this point that registration is a key issue while discussing the place of AI in the private law system.

Thirdly, it is possible to maintain that AI should not be endowed with legal capacity, because of moral or safety reasons; on the one hand, it should not keep any benefits, rights or property, arising as a result of the juristic acts in which it participates, because it is not a human or an organization of humans and it could use such assets to the detriment of human individuals or humanity as a whole. On the other hand, it cannot be burdened with any obligations because, at least today, it has no actual competences to be legally obliged. The consequence is that it cannot act on its own account or its own name. But despite this, according to this opinion, it can have the specific capacity to act on the account and on behalf of a legal subject, a

human or juristic person; as such, it may act as a mandate or agent. Of course, such a position seems to be difficult to defend when someone insists that the mandate or agency relationship demands that the entity which is the agent has legal capacity and that concluding a proper agreement between the agent and the principal is indispensable, as it is traditionally regulated in contemporary Western legal systems (cf. Article IV.D. – 1:101 of DCFR). But are these assumptions really imperative and logically necessary?

Despite their varied forms, algorithmic trading systems are usually recognized as an exception, which rarely makes jurists appreciate the need to award AI with the capacity for juridical acts of some kind and scope. Even when the conclusion of a juridical act results from the choice made by AI, this situation is regarded at most as if AI were the expert or advisor (not a slave as some try to insist[14]), and that it was the man or juridical person being party to the contract who concluded the juridical act by themselves, with the advice of the AI. As can be seen, especially in the case of high-frequency trading systems, such an assumption clearly flies in the face of the facts. The speed and mechanism of action of these systems are so great that no human is capable of controlling them in the course of their work.

Furthermore, it is often overlooked that many acts are regularly performed by natural and artificial entities which, although usually being classified as physical acts (i.e. ones of natural facts), have a strong legal dimension and significant legal effects. Therefore, the performance of such acts should not be entrusted to a random entity who is only physically able to perform them. Actually, in practice, it is usually required to have some kind of mandate for performing such physical acts on behalf of an authorized person; although, according to civil law rules, the mandate is given to the facilitation, negotiation or conclusion of a contract or other juridical act (DCFR, p. 559, also IV.D). No reasonable entrepreneur would send a mentally-disabled, although physically-strong and obedient person to collect a purchased computer from a shop if the shop is the place of performance of the sale agreement. Despite being seemingly only a carrier, the mentally-disabled person is not able to examine the computer to check whether it conforms to the contract or notify a lack of conformity within a reasonable time, which is necessary to keep certain legal claims (DCFR IV.A. – 4:301-302). This fact must be an important argument during the debate when it is well known that nowadays robots equipped with autonomous AI systems regularly perform deliveries, collections[15] or quality control of goods on a mass scale.[16]

Technological advances will soon limit the participation of humans in the economic sphere solely to the consumption of goods, because such goods will be produced, sold, bought, delivered, and checked by AI. As such, and if it is insisted

[14] Katz (2010), Pagallo (2011), Pagallo (2018), Corrales et al. (2018), Navon (2021).

[15] In China in 2020 many companies began using unmanned delivery robots and drones for delivering orders. Cf. Rui (2020), Snoeck (2020).

[16] The motor industry has long used robots in assembly lines and for product control—Deaton (2009).

that AI should lack the capacity for juristic acts, this will result in increasing numbers of practical problems. Therefore, there is a great need for further deliberations on the issue.

4.3 A Legal Capacity of AI and Its Capacity for Juridical Acts as a Function of Registration

It is quite common for AI systems to participate in legal transactions, and this demands some reaction from legislators. This state of affairs is indeed a threat to the axiological consistency of private law, which is based on the principle of equality of participants of legal transactions, with any exceptions being explicit and justified. In this case, the status is not equal: some participants are legally capable of making a declaration of intent in the transaction and others are not. However, it should be noticed that the principle of equality in private law is not both vertical, i.e., while dealing with the officers of the state, and horizontal, i.e., while dealing with other participants of juridical acts, as in public law; rather, it is understood only horizontally and situationally, in the sense that the parties to a concrete transaction are equal. Therefore, to satisfy this principle, it is not necessary for AI to have some general unconditional legal capacity and capacity for juridical acts of a certain scope; rather, it is important that the concrete AI which participates in a legal transaction is treated equally as the other party of the transaction. If this is not the case, this fact must be explicitly justified, as it happens when the parties are consumer and entrepreneur.

Considering all these difficulties, we are convinced that the best solution is to require an AI to be registered in a public register or certified in a publicly-authorized institution. Hereinafter, we will refer to such a procedure, whatever its shape, as *registration*. The registration of participants of legal transactions is not something unknown. Usually, it is required that juridical persons be registered; upon the moment of registration, these persons gain the legal capacity and capacity for juridical acts within a scope not contrary to the essence of a juridical person. As it was mentioned earlier, a juristic person cannot marry, adopt a child, testify, or be an author of a work, and so on, but can perform any act of a commercial character. Even if the register specifies the limits of a juristic person's activity, for instance, the pursuit of an international trade in exotic wood, exceeding the registered domain usually does not result in the nullity of the concluded juridical acts. Its effects are limited to internal relations; for example, members of the board may be liable in front of associates or funders.

In the case of AI, the registration should yield different effects: admittedly, the moment of registration would be critical for gaining legal capacity and the capacity for juristic acts, but the scope of these capacities should be limited to that described in the record of the register. Exceeding this scope should result in the voidance of the juristic act.

The issue of registration, as it was mentioned in other parts of the book, is a question of public, not a private law. It is not possible to predict how in the future this problem will be regulated. Although some proposals already exist in the Commission's Proposal 2021, their purpose is not to enable civil law institutions. They relate to high-risk AI and their two aims are robust monitoring and evaluation mechanism (Proposal 2021, 5.1. p. 12):

Article 6

Classification rules for high-risk AI systems

1. Irrespective of whether an AI system is placed on the market or put into service independently from the products referred to in points (a) and (b), that AI system shall be considered high-risk where both of the following conditions are fulfilled:
   (a) the AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonisation legislation listed in Annex II;
   (b) the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonisation legislation listed in Annex II.
2. In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall also be considered high-risk.

Article 51

Registration

Before placing on the market or putting into service a high-risk AI system referred to in Article 6(2), the provider or, where applicable, the authorised representative shall register that system in the EU database referred to in Article 60.

TITLE VII

EU DATABASE FOR STAND-ALONE HIGH-RISK AI SYSTEMS

Article 60

EU database for stand-alone high-risk AI systems

1. The Commission shall, in collaboration with the Member States, set up and maintain a EU database containing information referred to in paragraph 2 concerning high-risk AI systems referred to in Article 6(2) which are registered in accordance with Article 51.
2. The data listed in Annex VIII shall be entered into the EU database by the providers. The Commission shall provide them with technical and administrative support.
3. Information contained in the EU database shall be accessible to the public.
4. The EU database shall contain personal data only insofar as necessary for collecting and processing information in accordance with this Regulation.

That information shall include the names and contact details of natural persons who are responsible for registering the system and have the legal authority to represent the provider.
5. The Commission shall be the controller of the EU database. It shall also ensure to providers adequate technical and administrative support.

Besides, this proposal seems to be too pared down. It does meet the expectations of AI providers but does not admit that in fact all AI systems are potentially of high risk, each and every one of them, and all together. It is difficult, if not impossible, to separate the risks associated with the actions of a concrete AI system as a separate object (which may be an innocent or even a beneficial device) from those which are emergent on AI as a phenomenon in general. This problem is well understood by Stahl (2021, p. 49) who lists and describes 44 different issues and risks arising from the growing presence of AI in the world: lack of privacy, misuse of personal data, security problems, lack of quality data, lack of accuracy of data, problems of integrity, lack of accountability and liability, lack of transparency, bias and discrimination, lack of accuracy of predictive recommendations, lack of accuracy of non-individual recommendations, harm to physical integrity, disappearance of jobs, concentration of economic power, cost of innovation, contested ownership of data, negative impact on justice system, lack of access to public services, violation of fundamental rights of end-users, violation of fundamental human rights in supply chain, negative impact on vulnerable groups, unfairness, lack of access to and freedom of information, loss of human decision-making, loss of freedom and individual autonomy, unequal power relations, power asymmetries, negative impact on democracy, problems of control and use of data and systems, lack of informed consent, lack of trust, potential for military use, negative impact on health, reduction of human contact, negative impact on environment, unintended, unforeseeable adverse impacts, prioritization of the "wrong" problems, potential for criminal and malicious use, machine consciousness, "awakening" of AI, autonomous moral agents, super-intelligence, singularity and changes to human nature.

Therefore, it seems to be reasonable to register all used AI systems. We postulate that in the future this, or another, register should be used also for the purposes of determining which AI systems would have legal capacity and the capacity for juristic acts, and in which domains this would be valid. Such a register should be characterized by the following features:

1. The register should be global (it is not very probable) or regional; for example, it may cover the European Union.
2. The register should not be fragmented too much; if possible, it should be unified.
3. All types of AIs should be required to be registered, although the requirements concerning registered data for different types may be different. In the register, the legally-relevant types of AI should be classified, ranging from the forms which have no capacity for juristic acts, to those which are broadly capable of juristic acts.

4. The moment of registration should be the moment from which the registered AI would be allowed to act or would gain the capacity for juristic acts of a certain scope.
5. The register should include all the necessary information, including the scope of activity of the given AI, intended purposes, the technology used for its creation, the risk connected to its activity. In White Paper 2020 (5.D.b) it is postulated:

> Ensuring clear information to be provided as to the AI system's capabilities and limitations, in particular the purpose for which the systems are intended, the conditions under which they can be expected to function as intended and the expected level of accuracy in achieving the specified purpose. This information is important especially for deployers of the systems, but it may also be relevant to competent authorities and affected parties.

The register should be overt.

When acting in legal transactions, the AI should identify itself with a symbol, which would allow other participants of legal transactions to easily check any necessary data, such as the scope of registration, the validity of certificate and the person liable for the given AI.

The register should give a legally-binding guarantee that its records are true and offer protection for entities which act based on the information coming from the register.

The register may be linked to some mechanism intended to simplify the process of claiming damages, for instance, compulsory insurance.

The register should keep information and documents on the process of creating the AI. For instance, in White Paper 2020 (5.D.b) it is postulated that:

> [. . .] the regulatory framework could prescribe that the following should be kept:
> - accurate records regarding the data set used to train and test the AI systems, including a description of the main characteristics and how the data set was selected;
> - in certain justified cases, the data sets themselves;
> - documentation on the programming and training methodologies, processes and techniques used to build, test and validate the AI systems, including where relevant in respect of safety and avoiding bias that could lead to prohibited discrimination.

The register should generate a legal presumption, one parallel to the principle of reliability of land and mortgage register (i.e. a warranty of authenticity of land and mortgage register), that the AI registered as capable for juristic acts within a certain scope is actually capable within this scope. This presumption should be rebuttable in some specific circumstances. For example, if it were proved that this AI system has no UCD (cf. Sect. 3.2.1 *in fine*), because of a significant mistake in the system, incorrect data or hacking. If the presumption were rebutted, the juridical action concluded during a state proven to lack UCD should be acknowledged as void.

If an AI which acted in legal transactions were not registered or acted outside the registered range of activity, it should be acknowledged as illegal, it should be eliminated physically, and all its potential juridical acts should be void *ex lege*. Admittedly, other less drastic consequences are possible. For instance, the juridical acts actually concluded by such an AI could be acknowledged as directly performed by the user as if the AI were only his tool, or such an AI could be acknowledged as *falsus procurator* (an agent acting without or beyond the mandate). However,

accepting such solutions would not dismiss the problem, but would upset the certainty of the legal transactions. Without registration, or in transactions outside its scope, both the participants of the transactions and the user would lose the certainty of who is responsible for the AI, whose property it is and the mechanism of its action; there would also be no indication of its risk and whether it may be used by swindlers, among others.

Similar ideas, although limited, may be observed in Proposal 2021, for which a very important notion is "intended purpose". According to Article 3 of this proposal:

> (12) 'intended purpose' means the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation; [. . .]

Moreover, Annex VIII point 5 of Proposal 2021 includes a "description of the intended purpose of the AI system" among information to be submitted upon registration of high-risk AI systems; in addition, Article 71 authorizes Member states to lay down rules on high penalties for infringements of the proposed regulation and for supplying incorrect, incomplete or misleading information to notified bodies and national competent authorities. Furthermore, Article 65 lets the authorities of the Member State impose restrictive measures when dealing with AI systems presenting the risk, among them withdrawal of the product from the market, what means "any measure aimed at preventing the distribution, display and offer of an AI system" (Article 3 point 17).

Clear similarities can be seen between some Proposal 2021 solutions and our proposal, when considering that the intended purpose of an AI system designed to conclude certain juridical acts, such as mechanical or algorithmic trading systems, is to exactly conclude these juridical acts. If the registration of the intended purpose were acknowledged to be constitutive for the capacity for the juristic acts within the registered scope, the result would be the same. However, it should be stressed that our ideas are much more far reaching. We think that in a longer perspective the full registration concept would appear the most adequate.

The above proposal, despite its apparent size, could secure the market from a possible deluge of various opaque AIs; these would be true black boxes because there would be no trustworthy source which would give any reliable information about them.

4.4 Capacity for Juridical Acts by a Human User of AI

The growing use of AI could soon force a remodelling of the concept of capacity for juridical acts of humans. This is unavoidable because the development of novel technologies, including AI, gives humans new possibilities, measures, and ways of participating in legal transactions. As such, individuals in fact do not need to be conscious and free at the exact moment of concluding a juridical act: they can buy

and sell things while sleeping or being captured by terrorists; furthermore, they do not need to be educated enough to understand the mechanisms of the market, as the AI knows everything necessary to make them rich. In the traditional conceptual frame, they would be acknowledged as incapable of making a declaration of intent.

Two situations should be analysed:

(1) A human being fully capable of juristic acts starts the operation of AI system which constantly acts on his account, e.g., concludes agreements. After some time, this person stops being capable of juristic acts, although this state may not be formally acknowledged, at least until a certain moment, for example, the person is a victim of a road accident and is in a coma.
(2) A human who is permanently not capable of juristic acts from birth due to being a minor or mentally disabled, starts the operation of AI system which constantly acts on his account, e.g., concludes agreements.

Re. 1) A man buys a refrigerator, autonomous and connected to the web, whose function is to provide the owner with food. This is most likely a far more advanced refrigerator than any widely available on the market today, mainly due to the price of such technical solutions. It is capable of learning about the needs of the owner step by step based on past decisions and readings from sensors placed in the body of the owner and in his environment, and orders adequate quantities and kinds of food. One day the owner falls victim of a road accident and loses control over his body and mind: he is not aware of his needs, is not conscious of his actions, and does not understand the institution of money and contracts. However, the refrigerator orders the food according to the earlier preferences of the owner for some time. Then it learns that some of the products are not used at all, and completely different products are chosen. Hence, it changes the orders.

Alternatively, the owner of an internet business used for running an algorithmic trading system could fall victim to an assault and slip into a coma; however, his AI trading system would still continue to trade.

(a) When contracts are concluded by systems which are only automatons and are not capable of autonomous decision making, it is difficult to solve the problem of a human falling into a coma after starting an AI, only with the aim of protecting the other contractor's confidence and interests, particularly as, in such cases, the AI system only acts as a complex carrier of a human's will. This will, after all, may endure over time; therefore, according to generally-accepted rules, any loss of capacity by the addresser after sending the declaration of intent to the addressee should not cause the invalidation of this declaration. A similar example is presented by a man declaring his will concerning all contracts which could be concluded in the future: more specifically, a businessman who places a vending machine with refreshments in the commercial centre but has a stroke and falls into a coma while coming back to the office. In such a case, no one would say that the refreshments bought after his loss of consciousness are not the object of a valid transaction. A *prima facie* more complicated example is when an automatic, but not autonomous, AI concludes individual acts based on

some valid framework contract. Some would say this example is even simpler, because the agreements made by a vending machine may have different contractors, i.e., each bottle of refreshment may be bought by a different person, and the framework contract is concluded between stable parties which are then involved in all particular juridical acts under the framework contract. When the computer system is only an automaton, it is the will of its user that imposes the rules and parameters governing future transactions. As such, it is justified to claim that, despite the loss of consciousness by the user of a trading system, any later acts performed by this system are only accessory to the original will of a man, undertaken and expressed in a proper way; this would be particularly true if the framework contract determined a limited number of acts, or these acts were limited in the other way, for instance by date or worth. When the framework contract is open-ended, the solution of the problem may be to provide a clear interpretation of the declarations of intent which serve as its foundation. If these declarations were made by reasonable persons, it is not probable that, according to their will, they intended the contract to be eternal. Pleading the will of reasonable persons is justified by the rules of interpretation of various Western legal systems, also by the DCFR: "a contract is to be interpreted according to the common intention of the parties even if this differs from the literal meaning of the words" (II. – 8:101: 1) but "the contract is, however, to be interpreted according to the meaning which a reasonable person would give it if an intention cannot be established under the preceding paragraphs" (II. – 8:101:3:a).

After all, even if the parties were not far-sighted enough to appoint some way of termination of an open-ended contract, all contemporary legal systems allow for the possibility to do so. Even a life-annuity agreement ends with the death of the life-annuitant, which is a future but certain event. So, if the framework contract is open-ended, and the number of particular juridical acts under this contract is not determined beforehand, there are two possible positions that can be taken.

Firstly, in such a case, every particular act under the framework contract should be treated as a separate agreement, and each mental state of the party at the moment when he declares his intent, which is then only "carried" by non-autonomous AI, should be decisive for determining the validity or voidance of each agreement. However, upon closer analysis, it is unclear why the time-limited framework contract should be treated as a single juridical act presented "in episodes", whereas an open-ended framework contract acts as a kind of "clamp" linking separate juridical acts or "sequels". The mode of specifying the duration of the framework contract seems to have nothing to do with its unity or divisibility. Still, the key lies in the proper interpretation of the declaration of intent. It is not likely that any such contract would explicitly state that it expires or is suspended, and that any agreements concluded under it with the help of non-autonomous AI are void when the party using this AI becomes unconscious; this is not only because the parties are rarely so far-sighted, but because if they are reasonable, they would not want this state of affairs. It would not serve their interests: every period where the user of the

non-autonomous AI falls asleep, attends a boozy party or even suffers a short illness with a loss of consciousness would justify stopping transactions and require a commercial relationship to be built again from the beginning. Furthermore, it would not serve the interests of the other party, who would be dependent on the health and conduct of the contractor. No reasonable person would agree to such terms.

On the contrary, framework contracts are concluded to maintain the stability and persistence of the relationship. Therefore, any short lack of consciousness by a party who uses non-autonomous AI for concluding particular acts should not be a reason for questioning the transactions made according to the framework contract; this is also in line with the principle of interpretation in favour of a contract (II. – 8: 106 DCFR). Of course, by taking this point of reasoning, one could note the problem of differentiating between "short" and "long" or "lasting" loss. This objection, however, is not very significant. The courts decide such issues all the time when they declare a person incapacitated, or when they decide to delegate medical decisions about a person to a curator.

Therefore, summing up, the second position about the validity of transactions concluded under the framework contract by non-autonomous AI during the loss of mental health by one of the parties, should depend on the formulation of the framework contract; in addition, when a contract says nothing, the validity should depend on the fact of whether the loss is long-lasting. This reasoning may be supported by one more argument: public law most likely will demand high standards of technical security and human oversight from AI providers and users to ensure security, even if a person using an AI trading system loses control over their own mind.

(b) When an AI system is autonomous, the issue is more complicated, because such a system *ex definition* does not "carry", one-to-one, the will of the user. This situation may be at best interpreted as some anticipation of this will by the AI or rather the realization of the interests of the user by the will of the AI. Hence, it is necessary to carefully consider the AI's capacity for juristic acts. It seems obvious that when an AI acts autonomously, the capacity of the user to express his own will becomes beside the point when considering the validity of the concrete act concluded by the AI: after the moment of activating, it is not the user who decides, but the AI alone that choses between options. In the example given above, it is the refrigerator that makes decisions what products should be bought, from whom and for what price; it is really not important what the user thinks, despite him being the one who eats the products or makes the refrigerator work. Imagine buying such a fridge for old, infirm parents; if they have high blood sugar, the fridge will buy them products with a low glycaemic index, even if they would prefer some cake. Despite the joylessness of such an existence, it nevertheless shows that the AI would work efficiently, even if the beneficiary of the action could not formulate or express his will. This is consistent with the previously-described idea that AI should be treated in contractual relations as an authorized agent having certain capacity for juridical acts determined by its

registration. If the mandate was given by the principal in a proper way, by a person who was capable of doing so at the moment the mandate was given, then as long as the agent acts according to it, the capacity of the principal at the moment of a particular juridical act is not subject to examination. Therefore, if an AI system were registered and then activated by a person who is legally capable, any further action by the AI would not be burdened with the necessity of examining the capacity of the user each time. The authorization should be valid until the court or other competent entity acknowledges the lasting loss of competence of the user and decides to cancel the authorization.

Such a solution, despite some doubts, would not be dangerous for the interests of the principal. As noted above, the high standards of technical security and the human oversight demanded by the registration should be sufficient that when the circumstances change, e.g., the principal loses consciousness for an extended period, the AI system should quickly learn about this fact and adapt its decisions accordingly. If, for instance, the fridge makes decisions about the food based on the life parameters of the user (e.g., lifestyle, weight, blood pressure, cholesterol level, blood sugar level, iron content) it will quickly "notice" that the user is not conscious and will stop ordering food.

This way of thinking should not be treated as an axiom of course. The default principle should be that when the user dies, the AI loses its authorization to act on behalf and on account of the user. Although there could be some exceptions; for instance, when the AI system is a part of the business, the user could decide that AI would work after his death until the inheritor decides differently.

Re. 2) An entirely different situation concerns the activation of the AI by a person who lacks the capacity for juridical acts. In such a case, no valid authorization can take place, neither by the AI nor by an unsuitable human agent. As a rule, the juridical acts of an AI cannot reflect on the legal situation of the user. However, when the action of AI is beneficial for the person who lacks the capacity for juridical acts, there are possible different solutions of the ensuing problem.

(a) If AI is an automaton, the situation is analogous to the one in which the person who activated AI performs juridical acts by himself. In such a case, it is possible that some acts may be concluded without authorization; when, in certain circumstances, they are confirmed by the principal who gained capacity (for instance, a child who would gain the age of majority) or a statutory representative (e. g., parents), if such an institution is accepted in the given legal system.
(b) It seems that when AI is autonomous and acts within the range intended for it the situation is different and there is a possibility to use the institution of *negotiorum gestio*, i.e. management of another's affairs without mandate or, in another words, an agency of necessity (each legal system uses a different name).[17] In such a case the person who lacks the capacity for the juridical acts is a *dominus negotii* (principal) and AI is the *gestor* (agent or intervener).

[17] Cf. DCFR Book 5 *Benevolent intervention in another's affairs.*

However, still the confirmation of the principal who gained capacity or a statutory representative is needed and this confirmation should be made without undue delay. However, there are some unexplained issues which arise when the *negotiorum gestio* is to be applied. The first one is an issue of activation which seems to be contrary to the essence of the institution of *negotiorum gestio*. When this institution is applied to human beings the initiative of action belongs to the intervener. If it were applied to AI (intervener)—human (principal) relationship, the initiative in fact would belong to the principal. This causes two doubts. One of them is the question who should bear the risk of activation of AI by the unauthorised person, because when AI's intended purpose is performing juridical acts the person who has not capacity for juridical act within a certain scope is certainly such a person. Shouldn't AI be secured against the activation by such a person? If so, is the AI who should bear the consequences of the unwanted activity? Or should the person who activated AI be responsible for unwanted consequences, as the action of AI was a result of his initiative? The second doubt concerns the duties of the intervener (AI), such as duty to inform about intervention and seek consent for further acts. It should be undisputable that the duties of the intervener (AI), such as duty to inform about intervention and seek the consent for further acts or the after-intervening duty to report and account to the principal and hand over anything obtained as a result of the intervention, should be performed towards either the principal who gained capacity or a statutory representative, not towards the person who lacks capacity. However, if the addressee of these duties is to be a principal who gained capacity, they have no sense as this person activated AI by himself. Meanwhile, when the addressee of these duties is a statutory representative, how would AI learn who this person is and by what means could it contact this person? All these questions are so significant that they put in doubt the possibility of applying the subject institution.

4.5 Capacity for Juridical Acts by the Juridical Person Using the AI

The above part of our analysis relates to the situation when an AI acts in the name of, and on account of, a natural person. We now turn to the situation when AI acts for a juridical person. In such cases, two kinds of limitations, legal and actual, would overlap: those concerning the participation of juridical persons in legal transactions, and others concerning that of an AI in legal transactions. From this perspective, three theoretically-possible scenarios may be analysed:

1. The juridical person acts through an AI system with its own separate legal subjectivity, which is a status on a par with that of a juridical person (e-person).
2. The juridical person acts through an AI system which has its own separate legal subjectivity, and which acts as this juridical person's body (organ).

3. The juridical person acts through an AI which has no own separate legal subjectivity.

When a given legal system recognizes and acknowledges a new type of legal subject, which is AI and which may be, but does not have to be, named an *e-person*, and which has capacity for juridical acts performed in its own name and on its own account, the validity of this subject's acts will be evaluated from the perspective of this capacity and the rules governing such subjects. Furthermore, the capacity to act in the name or on the account of another entity depends on the rules concerning this e-person; whereas, the capacity of the principal is important only in the moment of issuing the mandate and for the range of authorization. Of course, a juridical person cannot award more power to an e-person than it has itself. But when the mandate is properly given, every juridical act of an e-person concluded in the name or on account of the juridical person should be valid, unless the juridical person loses its juridical personality, for instance as a result of liquidation.

In fact, the second situation, i.e., when a juridical person acts through its body (organ), which is an AI system having its own legal personality, does not differ very much from the first one. Everything depends on the rules governing juridical persons of a concrete kind, and on the rules governing e-persons of a concrete kind. If a given system accepts that some AI systems may act as bodies of some juridical persons, we do not see any general or essential obstacles to their use.

However, if the juridical person acts through an AI which has no own legal subjectivity, this entails a number of different possibilities which depend mainly on the type and status of AI.

Firstly, the AI may work only as a tool of a juridical person. This would be the case if the AI is not an automaton and is not autonomous, for example, its task is only to provide its user with appropriate gathered and selected knowledge, and the user chooses the appropriate action of the AI and confirms it.

Alternatively, if the AI is an automaton but not autonomous, it may conclude juridical acts as a carrier of a juridical person's will. However, in the domain of natural actions, i.e., those described above as giving some unobvious juridical effects, or juridical actions, it should be treated as a tool of a juridical person, and its actions and their results directly burden the account of the juridical person. For instance, if unattended cash registers in the shop receive money from clients, they are treated as tools, but the legal effects of such an action, ergo concluding the sale agreement, are directly attributed to the juridical person who is the owner of the shop.

Finally, if the AI is autonomous but has no own capacity for juridical actions within the needed range, it cannot conclude juridical acts in the name of a juridical person, or on its account, but it may perform natural actions, i.e. these described above as giving some unobvious juridical effects, or juridical actions whose effects will be counted on juridical person's account. However, in this situation, the directness of the attribution of these effects to the juridical person may be doubtful, since this situation seems more analogous to cases when a juridical person uses employees. For instance, an autonomous AI system may control the financial

situation of a juridical person's debtors and send them demands for payment when it assesses that their financial situation looks risky. Alternatively, an autonomous AI system may provide a subscription-based legal advisory service for the clients of a juridical person, in this case, the client may receive appropriate legal advice at any time of the day and night after giving a question and entering some data. Neither sending the demands for payment to the debtors nor giving legal advice in the implementation of the subscription agreement, are juridical acts and could be performed by an employee lacking any power of representation of the juridical person; however, both actions bring some important legal effects which will burden the account of the juridical person as if the actions were performed by a human employee.

The issue of capacity for juridical actions when juridical person uses AI is the most important for concluding contracts, and this is examined in the next chapter.

References

Books and Articles

Corrales M, Fenwick M, Forgó N (eds) (2018) Robotics, AI and the future of law. Springer, Singapore

Deaton JP (2009) How Automotive Quality Control Works. HowStuffWorks.com. 5 October 2009. https://auto.howstuffworks.com/under-the-hood/auto-manufacturing/automotive-quality-control.htm

Hage J (2011a) A model of juridical acts: part 1: the world of the law. Artif Intell Law 19:23–48. https://doi.org/10.1007/s10506-011-9105-4

Hage J (2011b) A model of juridical acts: part 2: the operation of the juridical acts. Artif Intell Law 19:49–73. https://doi.org/10.1007/s10506-011-9106-3

Hasan R (2021) 5 Best Automated Trading Softwares 2021. https://atozmarkets.com/news/5-best-automated-trading-softwares-2021/, last access on the 4th of August 2022

Katz A (2010) Intelligent agents and internet commerce in ancient Rome, society for computers and law. https://www.scl.org/articles/1095-intelligent-agents-and-internet-commerce-in-ancientrome, last access on the 4th of August 2022

Navon M (2021) The virtuous servant owner – a paradigm whose time has come (Again). Front Robot AI, 22 September 2021 | https://doi.org/10.3389/frobt.2021.715849

Pagallo U (2011) The adventures of Picciotto Roboto: AI and ethics in criminal law. In: Bissett A, Light A, Lauener A, Rogerson S, Ward Bynum T (eds) The social impact of social computing. Sheffield Hallam University, Sheffield, pp 349–355

Pagallo U (2018) Vital, Sophia, and Co. The quest for the legal personhood of robots. Information (Switzerland) 9(9):230. https://doi.org/10.3390/info9090230

Rui Z (2020) Meituan implements delivery robots and drones. China.org.cn. 29.11.2020. http://t.m.china.org.cn/convert/c_p8tCh0s7.html, last access on the 4th of August 2022

Smits JM (2014) Contract law. A comparative introduction. Edward Elgar, Cheltenham

Snoeck J (2020) Alibaba to start delivery with autonomous robots. Retail Detail. 22.09.2020. https://www.retaildetail.eu/en/news/general/alibaba-start-delivery-autonomous-robots, last access on the 4th of August 2022

Stahl BC (2021) Artificial intelligence for a better future: an ecosystem perspective on the ethics of AI and emerging digital technologies. Springer, Cham. https://doi.org/10.1007/978-3-030-69978-9

Varul P, Avi A, Kivisild (2004) Restrictions on active legal capacity. Juridica International IX/2004: 99-107. https://www.juridicainternational.eu/public/pdf/ji_2004_IX_99.pdf, last access on the 4th of August 2022

Wood G, Sutphen L (2015) FIA Guide to the Development and Operation of Automated Trading Systems, https://www.matbarofex.com.ar/documentos/mpi/fia-guide, last access on the 4th of August 2022

Documents

Convention on the Rights of Persons with Disabilities and Its Optional Protocol (A/RES/61/106) adopted on 13 December 2006 in New York, entered into force on 3 May 2008, https://www.un.org/disabilities/documents/convention/convention_accessible_pdf.pdf, last access on the 4th of August 2022

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), L 178/1, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031, last access on the 4th of August 2022

Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (recast) L 173/349 (MIFID II). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0065, last access on the 4th of August 2022

European Union Agency for Fundamental Rights (FRA). Legal capacity of persons with intellectual disabilities and persons with mental health problems, 26.7.2013. Vienna – Austria, https://fra.europa.eu/en/publication/2013/legal-capacity-persons-intellectual-disabilities-and-persons-mental-health-problems, last access on the 4th of August 2022

Recommendation No. R (99) 4 of the Committee of Ministers of the Council of Europe to Member States on Principles Concerning the Legal Protection of Incapable Adults (1999). https://www.coe.int/t/dg3/healthbioethic/texts_and_documents/Rec(99)4E.pdf, last access on the 4th of August 2022

Chapter 5
Consent

5.1 Introduction

The issue of concluding contracts by AI, or the use of AI, has been described many times from different perspectives. Generally, though, these all come down to the question of whether the fact that the contract is concluded in such a way requires legal systems to modify some private law rudiments or existing regulations, as well as these related to e-commerce or algorithmic trading systems. However, a frontal attack to this problem is probably not the most desirable strategy; a more effective one will be to try solving each particular problem separately, synthesizing the findings and then giving general rules.

As many times in this book, we start from an example of a practical operative case of an AI system whose main intended purpose is to conclude agreements. It should again be emphasized that although our analysis relates to weak AI, such an AI is nevertheless the most advanced type of system possible today; despite not being commonly accessible, it can nevertheless operate in the kind of social, legal or technological milieux which would allow it to use its power. Again, we will return to the case of the autonomous refrigerator; although it will exist within the IoT, its use will nevertheless generate the most typical legal problems.

The autonomous refrigerator is a device equipped with an AI system capable of analysing the past decisions, preferences, habits, and the everyday private and professional life of its user, as well as, of course, the content of refrigerator. It is also able to learn by itself and make conclusions about things the user should eat and place orders for these on the internet. This device not only stores the food in cool conditions and regularly restocks itself, but also makes decisions regarding what, when and where to buy supplies. It may also function as a cook, preparing the meals for the user. For this machine to be effective, it must have access to a wealth of data: all information about the family and friends of the user, as well as of the user himself, increases its effectiveness. This information is taken not only from its own sensors,

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_5

detectors, or cameras but also from a complete system of other sensors placed around the home and body of the user. The data about the food market is taken from the internet. All these information is collected and analysed constantly, and the fridge constantly learns not only what decisions will not duplicate the decisions of the user (milk is out—buy milk; it is Friday—buy the bottle of wine etc.) but also predicts what the user wants, or he would want if he made the decisions in a well-informed and free way. For example, it would notice that tomorrow is the user's, or his mother's, birthday and something should be prepared for the occasion. The user may not know it, but the AI could also make deductions by observing social media; perhaps if it knows the user is seeing his girlfriend that night, it may order a special type of French cheese that is popular in the social bubble where the user is active. However, such an advanced system connected to so many devices, computer and mobile apps, having access to the correspondence and conversations of the user will be capable of much more. For instance, it would be capable of constant analysis of the user's body functioning through sensors in smart watches, smart bands, chips, applications and cameras in smartphones; combined with the knowledge of everyday and social habits, this data would allow the AI to predict potential nutritional dysfunctions or what substances may be needed in real time. For instance, the user's diet may need to be modified when taking an additional job or starting karate lessons. AI will make contract decisions and will make the declarations of intent which do not mirror the content of the user's will but are rather the projection of the hypothetical will, which would never be so or not be so but for the decision of the fridge: the final will of the user is *de facto* restricted to confirming what was decided earlier by the system. The correlations perceived by the advanced and self-learning AI, which has access to a huge pile of data about the user, may be so complex and multilevel that they would not be consciously noticed by a human. This may well the equivalent to human intuition.

As it is commonly known, the efficiency of AI depends on the quantity and quality of data supplied to the system. The functioning of the refrigerator, which is here only an instance of any advanced AI system, would not be possible without at least basic data about the preferences of the user, e.g. what he has in his fridge, but true virtuosity would require access to all the possible information, including sensitive data concerning health and privacy. Of course, this would be possible only if the user decides to share this information with the AI and determines how far the system may go in analysing it, the extent to which it may draw conclusions, and the limits of the AI's "freedom".

This example can *prima facie* highlight a number of problems challenging civil law. For example:

(a) the issue of attributability i.e., whether and under what conditions the AI's action may be acknowledged as the user's declaration of intent or whether the AI's action may be acknowledged as independent of the user, but the results of which may be attributed to the user;
(b) if it is decided that the AI's declaration of intent is an independent act (i.e., AI is not only technical tool for carrying the user's will)—the issue of determining

whether, and under what conditions, it may be acknowledged that AI declares its will or has the intent to bring legal effects, what it means in the context of AI, and how to resolve the potential conflict between the will of AI and its user;

(c) issue of the form (procedures) of AI's action in legal transactions: explaining whether, and if so, how, this action should be subjected to special regulation: for instance, imposing informational obligations while concluding agreements with people (agreements AI to Person—A2P[1]) but also while concluding agreements with other AIs (AI to AI—A2A), or providing special protection of AI's contractors or users;

(d) the issue of defects of declaration of intent (vitiated consent or intention); this is a multi-layered question because the defects are highly-differentiated facts. Regarding AI, such doubts are connected to a lack of consciousness, freedom or the potential for error, or to fraud or unfair exploitation (i.e. a defect *sensu stricto*, exploitation of the dominant position on the market or excessive benefit or grossly unfair advantage in other circumstances), which seems to be very difficult to grasp when AI is involved.

5.2 Attributability

5.2.1 The Construct

When an algorithm based on some earlier plan performs actions according to the scheme "if A then B", it is quite easy to attribute the action to a human, or other legal subject, according to civil law. In such cases, the artificial agent can be regarded as only a tool or means of communication of the legal subject.[2] It is the subject who acts and communicates its will, and the advanced computer device only transfers it. The role of the AI may be completely passive, as in the case of a ballpen or a telephone, or it may be more active, for example, when the software transfers an earlier algorithmized will in a more sophisticated way; in this latter case, the artificial agent can be treated as a courier. This courier may even act by itself, within the limits imposed by very complex algorithms, on the stock market or while concluding

[1] Consequently, we use also analogical abbreviations: A2H—relationship between AI and human, B2C—relationship between business and consumer, A2A—relationship between two Ais, and so on.

[2] Cf. United Nations Convention on the Use of Electronic Communications in International Contracts (New York, 2005) Article 12. *Use of automated message systems for contract formation*: A contract formed by the interaction of an automated message system and a natural person, or by the interaction of automated message systems, shall not be denied validity or enforceability on the sole ground that no natural person reviewed or intervened in each of the individual actions carried out by the automated message systems or the resulting contract.

agreements on internet platforms. Such automated transactions have long been regulated by legislation.[3]

Examples of possible technologies which may be used for such automated transactions are blockchain and smart contracts.[4] These, together with other instruments and methods of making electronic declarations of intent, are based on the assumption that the legal subject bears full attributability of the "declarations" made by software. This assumption is a consequence of the fact that when the computer's action is automatic, no new "will" arises. Real decisiveness and volitionality, and ensuing unpredictability, are the characteristics of autonomous AI, rather than its automatic form.

The attribution of AI's declarations to the other entity is of an entirely normative character and does not reflect reality, because only a very general connection exists between the AI's declaration and the will of the other entity. In fact, the will of the user only begins the process: it is the reason for activating the AI and concerns the declarations which will be made later; their exact content is not—ex definition—possible to predict by the user. As a matter of fact, the user's will is blanket. The user not only accepts many actions which will be carried out later by the AI, but also gives the AI the competence to decide when and to whom the declaration should be performed, as well as its content, and to declare this will. Of course, this transfer of competence is limited, for example by domain: food stored in refrigerator, house-managing or certain kinds of business. Therefore, the problem is to find a special link connecting the results of the decision made by the AI and its action with the user (human or any other legal subject). It is worth noticing than the more advanced and autonomous AI is the deeper the divide between the real will of the user and the will of AI expressed in the content of juridical act is. So, in fact this relation is gradable and it demands more than two extreme solutions.

Returning to the example of the refrigerator (cf. Sects. 4.4 and 5.1), it is easy to imagine that while the user sleeps or may be completely drunk, the fridge may place new orders for products which are unknown to the user, but they fit his preferences, social group and economical and cultural situation. Is it then justified to say that the

[3] Cf. Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (recast) L 173/349 (MIFID II). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0065, last access on the 4th of August 2022.

[4] Smart contracts actually are neither AI nor contracts. As Filatova (2020) puts it: "By their essence smart contracts are software programmes or autonomous software agents that automatically bring about some specified action or execute other actions relating to digital assets in accordance with a set of pre-specified rules. Like most other software programmes, smart contracts contain so-called 'if/then' statements having a different level of complexity [. . .] What is also crucial about smart contracts is that they are coded transactions which are always represented in a programming language, not in a natural (human) one. Hence, their conditions (namely, what follows 'if') can be executed automatically without any human intervention: when the conditions in the code are met, the program triggers the required action [. . .]." The further remarks in the Chap. 9. This does not, of course, preclude AI from occurring alongside smart contract technology, enhancing its capabilities. Cf. Zou (2022, pp. 41–58).

user concluded the agreement with the alimentary supplier? Also, is it important that the user was sleeping while ordering, or that he did not know the ordered products and he would never order them on his own initiative? It would be absolutely counterfactual to state that the intent of concluding the contract belonged to the user and it was only transferred and revealed by the AI. Indeed, today it is not possible to create an AI which would exactly predict and realize the user's will.

In civil law doctrine, long before the rapid development of AI, a popular concept was that the connection between the action of artificial agent and its user may be described by three simple models.[5] While the simplest model is that of a tool, the most radical model awards AI with legal subjectivity. The latter model was proposed by Resolution 2017, but was rather abandoned by EU bodies. The third, half-way, model uses the institution of mandate, so the artificial agent is qualified according to its name as an agent who is artificial (artificial representative), which, in contrast to a "normal" agent, does not need to have legal subjectivity. This idea is promoted in this monograph in Sects. 4.3 and 4.4.

Some authors have proposed a model based on the use of the slave construct created in Roman law.[6] Briefly, the slave was generally treated as speaking tool (*instrumentum vocale*) and could not conclude agreements which bore obligations for its owner; however, it could conclude legally-valid agreements, with the permission or instruction of the owner, as part of a so-called *peculium*. A *peculium* was a piece of property allocated by the *pater familias* to his subordinate (slave or child), to whom he usually provided free administration (*libera* or *plena administratio*). Possession of the *peculium* did not change the legal status of the slave, who could own and possess nothing, but within the *peculium* he could enter into contracts without consulting the *pater*, requiring the *pater*'s consent or gaining his later confirmation. The *pater* was liable for debts burdening the *peculium* and could be sued by the creditor, but only to the value of the *peculium* (Frier and McGinn 2004, p. 263). However, the analogy here is rather distant, especially because according to the Roman law, a slave was a human being (*homo*), who although legally subordinated and not free, was not a beast or a machine. No one questioned his natural ability to be conscious and to act according to his own intent, although he could not use these natural abilities in legal transactions.[7] This is why even a Roman citizen could change his legal status several times during his life: he could be born free, then he could be enslaved and as a slave could be released. In contrast, an AI has not natural abilities at all, not to mention naturally understood consciousness or intent:

[5] Chopra and White (2011), p. 29; Dahiyat (2020); compare with the list of theories made by Habibzadeh (2016), and its synthesis presented in Sect. 4.2.

[6] Wein (1992), pp. 110–111, Kerr (1999), p. 237, Kerr (2001), Katz (2010), Pagallo (2010).

[7] Florentinus in his Institutes, book nine, wrote: Libertas ets naturalis facultas eius quod cuique facere libet, nisi si quid vi aut iure prohibetur. [. . .] Servitus est constitutio iuris gentium, qua quis dominio alieno contra naturam subicitur. [Freedom is the natural ability to do what one wishes, except if it is prevented by coercion or by the law. [. . .] Slavery is an institution of the law of nations, whereby, contrary to the nature, a person is made subject to another's ownership. Frier and McGinn (2004), p. 14.

all its actual abilities, exactly like its legal ones, are artificial. Applying the construct of the slave to AI may be treated only as a metaphor which should be used very carefully to avoid unjustified personification (anthropomorphisation, humanisation).[8]

However, the problem of the link between the action of an AI and the presence of the user in a legal transaction still exists. In our opinion, it is not possible in the long term to maintain the counterfactual legal fiction which regards AIs as mute tools in the hands of its user. This can only apply towards algorithms which act according to a given plan (if... then... scheme), and not to active, interpretative, autonomous, although still weak AI. Of course, giving AI full legal personality would solve many problems. Its action could be treated then like the action of other persons of the law, natural or juristic, also within the domain of acting in another's name or on their account. However, such changes do not seem necessary and could even be dangerous; therefore, as long as the legislative decisions were not so far advanced, it is possible to use the idea of limited, punctual legal subjectivity, as postulated in Chap. 2, and the existing concept of mandate or representation.

Here some parenthesis is needed. From the perspective of the theoretical model, it is possible to interpret the concept of *agency* or *representation* (different legal orders use different names) in two ways:

1. as a single contractual legal relationship under which a person, the agent, is authorised and instructed (mandated) by another person, the principal:
   (a) to conclude a juristic act (e.g., contract) in the name and on the account of the principal (a mandate for direct representation);
   (b) to conclude a juristic act (e.g., contract) in its own name, but on the account of the principal (a mandate for indirect representation);
   (c) to take steps which are meant to lead to, or facilitate, the conclusion of a contract between the principal and a third party.
2. as a legal relationship which consists of two elements:
   I. the authorization, empowerment (power of attorney) of a certain scope given by the principal to the representative (attorney-in-fact) to represent the principal (act in the name of the principal and on the principal's account) while concluding some juristic act and
   II. a legal or factual relationship between the representative (attorney-in-fact) and the principal being the foundation (justification) of the above authorisation, for instance, the contract of mandate, the contract of agency (in both latter cases the attorney-in-fact is called agent), employment contract, politeness, personal relationship of a certain kind, e.g., kinship.

The mode of regulation of the DCFR is closer to the first model, while, for example, Polish civil law is closer to the second. The second model seems to be more universal

[8] On the android fallacy and anthropomorphic rhetoric *cf.* Avila Negri (2021).

and, as will be clarified below, more useful for the purposes of solving the problems indicated in this book.

According to the above arguments (see Chap. 2), bestowing legal subjectivity does not have to be equivalent to giving full legal capacity and capacity for juridical acts. Indeed, legal subjectivity exists more as a function of the social role of a given entity; when this role is narrow, legal subjectivity may be punctual and embrace only a strictly-limited sector of reality. For example, it may only constitute only a single social relationship (state of affairs), if it is possible and expedient to do so. This is also the case when an AI is required to issue concluding agreements on behalf of its user. If it is accepted that the AI acts as an agent or representative of the user, this would entail accepting that, at this exact point, the AI is a legal subject; this may acknowledge the link between agent, i.e. the AI, and principal, i.e. the user. An AI acting in the name of, and on the account of, another person has the capacity for juristic acts in this narrow domain; however, according to the assumptions founding this book, this capacity is only valid when the required "job" is consistent with its aim (i.e. intended purpose), appointed in the appropriate register. However, it should be emphasised that this legal status does not mean that AI is a full person of the law which can act in its own name and in its own interest. While AI is not conscious, it should be an axiom that it has not any own interest, except for the interest understood as realizing its aim (i.e. intended purpose); in fact, its existence is always subordinated to realizing the more general human interest.

Such a solution is possible today without significant legislative changes. The binding rules of civil law do not exclude such an interpretation, especially if we take the second model of agency (representation) given above. Not only do most legal provisions not explicitly demand that the agent must be a full person of the law, although this is assumed as obvious, but also that there are no over-riding systemic arguments for being a full person in the eyes of the law. So far it has seemed obvious that the agent (representative, attorney-in-fact) has to be a full person of the law because there were no other candidates, i.e., natural persons or juristic persons, who would have the factual abilities to represent another person, to act in another person's name and on another person's account. Neither things, nor animals could do it. Now this state of affairs has been upset by the arrival of a new kind of entity. AI systems have a factual ability to act in another person's name and on another person's account, despite this ability not being natural. Why not let them do so?

Firstly, some say that such a construction is not proper because there should be an agreement concluded between the agent and the principal. As it was shown above, this argument is very weak, especially if the second model of agency (representation) is accepted. It is sufficient that some relationship exists, not necessarily a legal one, between the agent (representative) and the principal, which on the one hand justifies the empowerment of the agent by the principal, while on the other, obliges the agent to respect the instructions of the principal. In the case of AI, it may be accepted that the foundation of these relations is either the copyright to the AI, or the licence agreement based on which the licensee may use the AI (i.e. the user). Both elevate the position of the principal above that of the AI and subordinate the AI with regard to the principal.

Secondly, no one is endangered by letting AI represent legal subjects, even if the AI itself is not a full person of the law. The principal authorises the agent to influence the legal relationships within the domain of the principal's interests. For the principal, it is important only that the agent can conclude juristic acts and not that the agent has the legal capability to do so. Since, for the principal, it is not important whether the agent can be the subject of rights or obligations, as these exist in the domain of the agent's own interests. Legal capability is only necessary for a person who gives the empowerment to act, with the intended result being for himself. The agent does not have to be a full person of the law to play his role efficiently, he must merely have the capacity for acting as an agent. For the third party, the concluded acts do not bear any risk, because the principal is liable to him, not the agent.

Even if, despite the above arguments, the proposed interpretation were acknowledged *de lege lata* unacceptable, only relatively minor normative changes would be needed to favour this solution. Two directions are possible: (1) new rules could be drawn up to explicitly state that the agent does not have to be a person of the law, or (2) these new rules would create a special institution of artificial agent. In both cases, the rules should explicitly indicate the nature of the entity and the conditions under which it could perform this function. We are convinced that to ensure the safety of legal transactions, one such condition should be that the AI is registered under the aim (i.e. intended purpose) of representing persons of the law in legal transactions and has the UCD (cf. Sect. 3.2.1).

Accepting AI as a legally-acting agent could resolve some of the problems given above, e.g. the problem of attributability. As stated by the DCFR:

II. – 6:105: When the representative acts:

(a) in the name of a principal or otherwise in such a way as to indicate to the third party an intention to affect the legal position of a principal; and

(b) within the scope of the representative's authority, the act affects the legal position of the principal in relation to the third party as if it had been done by the principal. It does not as such give rise to any legal relation between the representative and the third party.

II. – 6:104: (1) The scope of the representative's authority is determined by the grant.

This scheme effectively regulates this phenomenon: it is not the user who decides and declares his will, the AI does so, but in his name and on his account. Therefore, if the AI acts within the scope of its authorization, its acts are binding for the user. If the AI, acting as an agent, strays outside this scope, its acts should be evaluated according to the rules concerning *falsus procurator*, as it is explained in the next part of this chapter.

For example, if the refrigerator buys food which is needed, according to its calculations, the legal transaction is valid. Although the user for whom it works does not make the decision to purchase the food and may not even know about it, the legal effects of the transaction go on his account, and burden him. However, if the fridge goes outside the scope of its responsibilities, for instance, it buys a car, this action would not have any effect on the rights and duties of its user.

Of course, for the sake of safety it must be demanded that another party of the transaction should be informed before that the contract is concluded by an autonomous AI acting as an agent of the user. Indeed, the proposed regulations go in this direction (cf. Proposal 2021)

5.2.2 Exceeding the Scope of Authorization

The admissibility of an AI acting in the name of, and on the account of, legal subjects does not mean that any such concrete action of AI would be proper and valid or that it would be attributed to the user. The example given above of the refrigerator which buys a car instead of food is an extreme one. A more probable instance is that an e-assistant too broadly interprets the command of the user and buys a car while the user asked it only to research the prices and models for later consideration. May such event be described within the existing legal institutions? It seems that it may. Returning to the concept of *falsus procurator* given in the DCFR:

II. – 6:107: Person purporting to act as representative but not having authority

(1) When a person acts in the name of a principal or otherwise in such a way as to indicate to the third party an intention to affect the legal position of a principal but acts without authority, the act does not affect the legal position of the purported principal or, save as provided in paragraph (2), give rise to legal relations between the unauthorized person and the third party.
(2) Failing ratification by the purported principal, the person is liable to pay the third party such damages as will place the third party in the same position as if the person had acted with authority.
(3) Paragraph (2) does not apply if the third party knew or could reasonably be expected to have known of the lack of authority.

II. – 6:111: Ratification

(1) Where a person purports to act as a representative but acts without authority, the purported principal may ratify the act.
(2) Upon ratification, the act is considered as having been done with authority, without prejudice to the rights of other persons.
(3) The third party who knows that an act was done without authority may by notice to the purported principal specify a reasonable period of time for ratification. If the act is not ratified within that period ratification is no longer possible.

Such rules considering persons purporting to act as representatives but not having the authority to do so could be applied to the actions of AI when it acts without authority or outside its boundaries; however, these rules require systemic correction with regard to property, as proposed in Chap. 8. Adopting the rule that the an action performed by the AI without, or outside, authorization is not attributable to the user,

does not bind him nor make the structure of the rules more problematic; however, it bears some complications of a practical nature. Such complications do not arise at all if the purported principal, i.e. the user, ratifies the AI's act and takes the consequences upon himself; alternatively, they do not occur when the user does not accept the consequences but the AI is legally and actually capable to be the party of the relationship, because then, the act gives rise to a legal relationship between an unauthorized AI and the third party. However, when no legal relationship exists between the purported principal or AI and the contractor due to various reasons, the AI itself has no legal capacity for being a party of the agreement of the given kind, for example, or the relationship cannot be continued due to the AI having no actual capability to perform its obligations under this relationship, some harmful consequences for the contractor may arise. In such cases, who would be required to offer compensation? Who should satisfy the claims of the contractor? There are several possibilities to be chosen by the lawmaker or the doctrine:

(1) the claims may be satisfied with the AI's own assets;
(2) the claims may be satisfied by the insurer;
(3) the claims may be satisfied by the person on whose behalf AI acted, or by the other persons, for instance the provider or producer.

It is important to emphasize here that the current analysis relates to the cases of an AI which has been registered and acts within the scope of the registered aim (intended purpose), i.e. representing other entities while performing juridical acts. The consequences of such action by an unregistered AI or its action outside its registered aim (intended purpose) will be examined separately.

Ad. 1. Making an AI liable for its own action is reasonable only when the AI has some property with which it could satisfy the claims. As it was argued, such a solution is possible, at least in some cases, and probably could turn out to be functional. An AI acting within some domains could be awarded the limited legal subjectivity needed to realize its purposes, such as representing some other legal subject, and to entitle it ownership of some assets (money). These assets, on the one hand, could help it realize its aims, such as paying deposits in the public contract system or guaranteeing participation in auction sales, while on the other hand, they could be used for satisfying the claims of third persons.

Ad. 2. Similar effects would be achieved by connecting the liability of AI to that of an insurer. The source could be either normal commercial insurance or a special fund (public or private) created solely for the needs of emerging technologies. In this latter case, various methods of financing could be used: the premium may be paid by the users, operators, providers, or even by AI itself. In the case that the premium is paid with the AI's own money, the above proposal of giving AI some property is still valid. Of course, the construction of the fund may be completely different: it may be financed by taxes taken from the users, or the producers or providers of AI. Also, different instruments may be used simultaneously, and may have more than one function (see Chap. 6 regarding AI authorship and copyright to works by the AI).

Ad. 3. Any potential liability of the user or other subject, such as producer or provider, for damage caused by AI's action without or outside authorization does not

fit in the concept of *falsus procurator* and it should be placed within the realm of tort law. Such liability could burden different entities, and the choice depends on the lawmaker. If AI is treated just as a regular product, and the focus is kept on the fact that it is the provider who is liable for the product and who should assure the proper operation and safety of AI, the liability for damage caused by the action of an AI without, or outside, its authorization would burden the provider. Such liability could be either fault or strict liability, although today's standard is strict liability. For practical reasons, strict liability or fault liability with the presumption of the user's (or other subject's) fault seems to be more effective than normal fault liability, especially since it could be bound to the compulsory insurance mentioned earlier. So if our hypothetical refrigerator, after a command "buy meat and dairy", bought a dairy (e.g. factory of yogurt) instead of dairy products (e.g. yogurt, butter, milk or cottage cheese), and the user did not confirm this contract because his business was only IT, the liability for damage made to a contractor by the invalidity of agreement would have to be strict liability or, perhaps, fault liability with the presumption of the user's (or other subject's) fault: if the AI had acted without authorization, it would be difficult to attribute the fault to anyone.

5.2.3 Acting Outside Registration

A different kind of problem arises when AI has no legal capacity for representing other entities while concluding juridical acts within the given domain. This may happen if an AI is not registered at all, or when the registered aim (intended purpose) of the AI does not embrace the representation of other entities while concluding juridical acts within the given domain (cf. Sect. 4.2). Acting outside the limits specified by the registration (as a matter of fact, outside the borders of the AI's legal subjectivity and capacity for juridical acts) should be acknowledged void, especially if the register complied with the proposed requirements: it is overt, easy to check, and guaranteed to mirror the true state of affairs. As it may be predicted the register would be completely digital and would allow both humans and AIs to see its records.

Any action by the AI outside the boundaries of its registration should be acknowledged illegal and not result in any legal effects, except for the legal (criminal, civil and administrative) responsibility of the person who placed this AI on the market or put it into service. However, in the case that the person instructing the AI, i.e. in whose name or on whose account the AI acted, knew about the action and gave his consent to carry it out, there are possible two conclusions:

(1) the AI's action is void and legally ineffective and the user (the actual principal) is liable according to the tort rules;
(2) the unregistered AI (or the AI acting outside its registration) may be acknowledged a mere tool which is only used by its user (its actual principal) and then the

consequences (also legal effects) of the AI's action lie on the user, also according to contract law.

Either model, or both, could be introduced to the legal system.

5.2.4 *Acting as a Legal Person's Body*

The next situation demanding examination is when AI is a member of a body (organ) or is a body (organ) of a legal person, e.g., the member of a board of the limited liability partnership. [9] Here, two elements should be taken into consideration: first, that the possibility to construct such a legal person's body is allowed by the given legal system, and second, if a given legal system accepts an AI as a member of a body (organ) of a legal person, then it is necessary to define the mechanism by which the AI's action may be attributed to a legal person.

This two-tier structure of the problem is nothing new. It can be observed in the discussion whether a legal person may be by itself the management organ of another legal person. Such a discussion is taking place in the European Union, which includes the German and French normative models. According to the German model, accepted e.g. in Poland, among other countries, only natural persons are allowed to hold the function of a legal person's organ; however, according to the French model, accepted by, *inter alia* Spain, Belgium and Czech Republic, legal persons may also hold the function of organs of another legal person.[10] This latter model includes two conditions: that the legal person functioning as an organ appoints a natural person who acts as its representative, and that this natural person is jointly liable because of his position. As a rule, European law is not in favour of either of these models: Article 47.1 of the Council Regulation (EC) No 2157/2001 of 8 October 2001 on the Statute for a European company (SE)[11] says that an SE (Societas Europea) statutes may permit a company or other legal entity to be a member of one of its organs, provided that the law applicable to public limited-liability companies in the Member State where the SE's registered office is situated does not provide otherwise. The law of the United States is not uniform in this regard, although the rule that only natural persons may be directors is predominant.[12]

The main reasons justifying the German model are the safety it provides for third persons and the clarity of the structure of the legal person. The French model is promoted as equally safe or more so, as the assets of the legal persons liable for

[9] The level of autonomy and importance of AI as a board member can vary widely, cf. Drukarch and Fosch-Villaronga (2022).

[10] del Val Talens (2017), p. 610.

[11] Council Regulation (EC) No 2157/2001 of 8 October 2001 on the Statute for a European company (SE) with amendments, L294/1, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32001R2157, last access on the 4th of August 2022.

[12] del Val Talens (2017), p. 615.

damages caused by the wrong management may be even more valuable than these of the natural person; it also does not endanger the clarity of structure in cases where overt registers are kept of legal persons, and is believed to be more flexible and effective from a business perspective. This approach, i.e. the French model, is also regarded as being better than gaining results by means of so-called management agreements. Indeed, concluding such agreements by legal persons with companies which act as management services runs the risk of a charge of abdication of authority by the board.[13]

As it noted above, it is also necessary to agree the mechanism of attribution of the action of the organ to a legal person. This depends on which the general theory of the legal person is explicitly or tacitly accepted. In the European tradition, two of many such theories were dominant, the first being the *fiction theory,* and the second being a realist theory called *the organic theory*. The fiction theory indicates that although a legal person has many real, or *material* faces, their presence only exists as a fiction in legal transactions, because this legal person does not have free will. The organic theory assumes that a legal person is a real entity, a social organism which is a "wholeness", and a part of this wholeness are its organs.[14]

These two theories differ as to their understanding of the mechanism of attributing the action to the legal person. According to the fiction theory, an action is attributed to a legal person in a similar way as to natural persons with no capacity for juridical acts (minors or incapacitated adults), and for whom the acts are performed by the statutory representatives. These representatives, which as natural persons have their own free will and capacity, act in the name of these persons and on their account. In contrast, the organic theory assumes that the organs of the legal person are an integral part of the wholeness of this person and are analogous functionally to the biological parts, i.e. organs, of a human being, by which and thanks to which, it can reveal its will. It is important to note that both theories are compatible, insofar that actions performed by the legal person's organs are made by these organs, and only the results are attributed to the legal person. However, the difference lies in the observations that according to the fiction theory, the organ acts from outside the legal person, while according to the organic theory, the organ is a part of its structure, i.e. it acts from the inside.

In recent times, a new concept in which a legal person is not capable to legally act on its own has gained popularity in Europe; as a result, if it is to act it must happen by the action of its organs, which are able to create the will of this person. The organs are not external to the legal person in whose name they act, although they are not the part of a single indissoluble wholeness, as postulated in organic theory. They just serve as legal organizational component parts of a legal person, and their will and action are attributed directly to the legal person by imputation (*Zurechnung*). Such

[13] del Val Talens (2017), pp. 613–622.

[14] Beran (2020).

organizational representation is treated nowadays as a third form of representation, next to statutory representation and representation based on authorization.[15]

The above considerations relating to the status of legal persons as organs of other legal persons are very useful when researching the problem of AI as a potential organ of a legal person. Upon further reflection, it is clear that there are no logical obstacles to allowing AI to hold the function of the organ of a legal person and there is no reason why the results of an AI's action performed "in the hat" of the organ cannot be imputed to this legal person.[16] This does of course assume that some concept of legal subjectivity, will and capacity for juridical acts for AI is accepted in the given legal system. However, as we have demonstrated in the other parts of this book (Chap. 2), it is only a matter of the political will of the lawmaker, which should be made after serious deliberation.

Of course, for practical reasons, it is difficult today to imagine that AI could be the only organ managing the legal person, although as a matter of fact, this depends on the tasks of this legal person. Therefore, it may be necessary to introduce the condition in law that at least one member of the board must be a natural person; such conditions are already present in some legal orders, for instance, the UK. [17]

5.2.5 AI-Representative Acting in Its Own Name

DCFR defines a situation when a person acts in its own name while being a representative of the other person thus:

> **II. – 6:106: Representative acting in own name**
> When the representative, despite having authority, does an act in the representative's own name or otherwise in such a way as not to indicate to the third party an intention to affect the legal position of a principal, the act affects the legal position of the representative in relation to the third party as if done by the representative in a personal capacity. It does not as such affect the legal position of the principal

[15] Mucha-Kujawa (2017).

[16] In Hong Kong the company Deep Knowledge Ventures appointed as a *de facto* member of the board of directors the robot Vital 2.0. Burridge (2017) Marc Benioff CEO of the company Salesforce uses the AI engine Einstein and let it attend the weekly meetings of executives. Cf. Bort (2017).

[17] Section 155 of the UK Companies Act 2006:

> 155. Companies required to have at least one director who is a natural person
>
> (1) A company must have at least one director who is a natural person.
> (2) This requirement is met if the office of director is held by a natural person as a corporation sole or otherwise by virtue of an office.

https://www.legislation.gov.uk/ukpga/2006/46/section/155, last access on the 4th of August 2022.

in relation to the third party unless this is specifically provided for by any rule of law.

This regulation is dictated by the reasons of safety and protection of the third party. It, therefore, clarifies who are the parties of a given agreement. If the third party does not know that a person with whom he concludes the agreement is somebody else's representative, or that this person acts in the name of somebody else, he should not be surprised by the fact that his contractor is not the person he interacted or negotiated with.

When an AI acts in legal transactions, the problem should not be so serious, especially because all proposed legal regulations concerning AI demand that an AI should inform the human interlocutor about it not being human. For example, Article 52.1 of Proposal 2021 says:

> Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.

Therefore, all parties involved in concluding an agreement should be conscious that they are talking and co-operating with an AI system. In addition, having the access to the proper register, they should know the scope of actions allowed by AI. If the AI is allowed only to represent other legal subjects, it is not capable of concluding agreements in its own name; if it attempts to do so, the contract shall be void and the liability will burden the user or provider, or any other person responsible for controlling its actions. However, if the given legal system awards the AI a wider scope of legal subjectivity, allowing it to act in its own name in some domains, than the situation should look like the one described in DCFR and cited above. It should be assumed that allowing AI to act in its own name would be associated with giving it some property and providing that it is the subject of compulsory insurance, which would be recorded in a proper register.

5.3 AI's Intent and Declaration of Intent

When the legal system allows AI to make an efficient declaration of intent, either in its own name or in that of the user , it should be clear what this "intent" means. This need seems obvious considering Article II. – 4:101 and Article II. – 4.102 of DCFR, which are representative of many EU Member States' regulation of civil law:

> **II. – 4:101: Requirements for the conclusion of a contract**
> A contract is concluded, without any further requirement, if the parties:

(a) intend to enter into a binding legal relationship or bring about some other legal effect; and
(b) reach a sufficient agreement.

II. – 4:102: How intention is determined
The intention of a party to enter into a binding legal relationship or bring about some other legal effect is to be determined from the party's statements or conduct as they were reasonably understood by the other party.

As it can be concluded from such words as "statements", "conduct", or "reasonably", these provisions demand that the intention should be understood based on its external manifestations, and not on the subjective attitude, emotions or consciousness of the entity entering in the legal relationship. The content of such external manifestations (statements, conduct) may be called the *declaration of intent*. By shaping of the concept of intent this way instead of subjective concepts, it is easier to include AI in legal transactions, because contemporary AI is not conscious and does not have any own private attitudes or emotions. Such concept of intent is proof that on the grounds of contract law, it is not the psychological internal will that counts, but the will understood as social communication fact.[18] From this perspective, the fact that AI has no psychological internal will becomes irrelevant. Therefore, whether an AI concludes an agreement in natural language or in machine language, or some other higher-level language understandable by the receiver, either directly or indirectly thanks to translation, for example, when the refrigerator orders some drinks, it is enough to acknowledge that it has an intention of concluding the agreement. Of course, the meaning of the conduct must be read in the social context of the given interpretative community[19]. More about the interpretation of the declarations of intent may be found in Sect. 9.3.

In establishing the content of the intent, there are of course some elements which are not expressed directly in the declaration of intent. But these elements can be objectivized too, for example, the knowledge of the person who acts with the intention of entering into the legal relationship; this is particularly the case if an AI acts as a representative (agent). As Lewaszkiewicz-Petrykowska (1983), p. 21 notes:

> In every case when arising of a concrete legal effect depends on the unawareness of the party, on the lack of knowledge of a given state of affairs, knowing this state of affairs by the principal or an agent excludes coming into being the effect, which is dependent on unawareness.

Therefore, if an AI collects some information and uses it to perform a juridical act in the name of another entity, it should be assumed that this information is also

[18] The question of intention of AI in the context of contracting is approached differently by Linarelli (2022), who links such a possibility only to much more highly developed AI.

[19] The good example of the social contextuality of the meaning of conduct may be the custom taking place in old times on the horse markets in Poland, where the agreements were concluded by the contractors spitting in their own hands and affixing this hand to that of the other contractor. No words were necessary.

possessed by the AI's principal. For instance, let's assume a case where the legal effectiveness of acquiring a property depends on the fact that the acquirer knows that the property has a defect (e.g., it belongs to a third person or is a subject of security), but the acquirer has no knowledge about the property and uses an AI system as an agent; if this AI gathered the data and learned about the defect while determining whether the act of acquisition is efficient, it should be assumed that the AI's principal also has this knowledge. Of course, this assignment here is of a purely normative nature and is independent of the actual flow of information between the AI and the person using it. It is a different matter, however, that systems used to act on behalf of and for the account of others should be constructed in such a way that the principal receives such information; the scope of this obligation, however, would already result from the specifics of the system and the contract between its provider and the user.

Because of their specific character, the legal transactions performed by means of AI should be examined more from the perspective of their form and applied procedures than from the perspective of their content. So, due to the need to protect the third party and the user itself, the law should demand (1) that the AI provides information that it is acting in its own name or on behalf of some other entity, (2) information on the limits within which AI can act, and possibly (3) giving special rights to the third party which concludes the agreement with the AI. This represents the direction of today's proposals for new legislation.

It is strongly justified that the contractor should know that the party of the juridical act is an AI system and not a human. Firstly, it could be considered a matter of fundamental rights and constitutional rules: it is a value of dignity which demands that a person knows that the interactions in which he participates are not interactions with people. Secondly, from a civil law perspective, the principle of contractual freedom requires that a person consciously enters a contractual relationship by consciously choosing the other party. Also, from a civil law perspective, such transparency is required by the principle of equality of civil relationship parties, because on the one hand, the AI may be much more powerful than a human contractor, by dint of its greater access to information and processing speed, and on the other, AI would know that the other party is a human, what certainly gives it an advantage. Hence, as it was mentioned before, Proposal 2021 (article 52.1) commands providers to ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system.

The above issue is associated with the problem of setting clear limits for the autonomous actions of an AI. As strong AI is yet to exist at the time of writing, and existing systems are usually one-task software with a narrow scope of activity, it is not currently possible that a refrigerator could buy a car or order a trip abroad. However, the potential of AIs is growing and future products, particularly those from the most powerful providers or which are platforms for more specialised applications, will be considerably more able. Therefore, there is a need to establish some general limit for the participation of AIs in legal transactions, prohibiting people from leaving AI property in a will for example, which would establish a general

standard legal capacity for AI. In addition, the scope of competence and capacity for juridical acts given to a concrete AI system should be kept in a register and be accessible to read from the system itself. Without such systemic securities, there is a risk of serious problems caused by AI acting illegally of outside its authorisation.

Considering the new risks faced by AI contractors who are not using AI themselves and the new vantages offered for AI users and principals, there is a pressing need for them to be awarded special rights. It is important to note here that the risks and possibilities associated with the AI – human relationship (A2H) are not the same as those of the business – consumer relationship (B2C). This results from several objective circumstances:

(1) intellectual vantage—the AI itself and its users would benefit not only from having access to an enormous amount of data and the incredible processing speed of the AI, but also from the fact that an AI does not need to rest, what means it can work permanently and does not yield to the emotions and suasions of the unconscious part of the human psyche;
(2) the ontological difference between humans and AI also entails considerable cognitive differences between the two—AI and humans have incommensurable differences regarding how they perceive reality and make decisions;
(3) the psychological oddity of the contact between a human and a machine, with humans consorting with a non-human source of will. This does not have any precedent in the history of contracts, because humans always concluded contracts with another humans, even if the party was a legal person. Now, a party to the contract, or the entity who undertakes the actions striving to conclude the contract is a machine.

All the above circumstances may cause inexpedient effects for the human participant, because human consent may be not free enough.

Of course, the institutions and mechanism characteristic of consumer law may be used. For example, the right to withdraw from a contract concluded with an AI for some specific period of time without giving a reason; upon such reflection, the human party may feel that the stimulus for concluding the agreement was down to persuasive techniques, informational imbalance or the very fact of communication with a machine. However, this right should not be limited to consumers only and to contracts concluded by real-time distance communication. Furthermore, this right of withdrawal should of course be available if the informational duty on the AI as a contractor were to be infringed.

There are also other mechanisms which could also be used. For instance, it can be ensured that the human contractor is aware that he is interacting with an AI and understands the consequences of such interaction; in addition, the burden of proof relating to facts significant for the interpretation and performance of the agreement can be attributed in certain ways. However, the practice of past years indicates that such mechanisms used to verify awareness may not be efficient; if commonly used, they may just be regarded by the recipient in the same light as the "informed consent" that most people click through when installing software.

5.4 Contracts A2A (AI-to-AI)

Another circumstance which should be taken into consideration is that AI systems may be placed on both sides of the agreement. Such situations will arise increasingly often because concluding the contracts by autonomous systems without any participation by humans has many advantages deriving from its significantly greater economic effectiveness: this offers the advantage of a shorter and faster process which can be concluded at any time of the day, and possibly at lower transactional costs. For instance, when the AI which manages the refrigerator discussed earlier, and thus the whole household and the life of the user, makes certain offers to purchase food, the best (optimal), most economical results would be gained when the other AIs used by shops or suppliers communicate on-line with it and conclude or coordinate the best contract. If such systems are advanced enough, they will always be better than any human. This is especially true for business (B2B contracts) where subjective elements, mirroring personal beliefs and preferences are less important.

The A2A contracts are different from H2A contracts at least in two respects:

(1) there are no reasons to protect one of the parties;
(2) the information is valid only when it is possible to be analysed and verified by AI, so the content and the form of information may be different in cases where at least one party is a human.

Therefore, because both parties are in the same situation, there is no need for the party to have any right of withdrawal due to the contractor being an AI. Also, the scope and the means of data exchange should be different, because AIs can communicate "in the same language". Therefore, H2A transactions and A2A transactions should be regulated by different legal rules, although the general principles applied to AI, for instance, that AI may act only within the scope limited by its registration, and the general institutions of the civil law, for instance, remedies on vitiated consent or intention or exceeding the scope of authorization by the agent are maintained.

5.5 Defects in the Declaration of Intent. Vitiated Consent or Intention

5.5.1 The Concept

The acceptance that AI is not only a tool which carries the will of the user, but that it also significantly complements this will or even substitutes it, forces a new approach to the issue of defects in the declaration of intent made in these new circumstances. The question is whether the facts regarding the user or the AI, or some other facts, are

important for the evaluation of the correctness or defectiveness of the declaration of intent made by an AI.[20]

For instance, after a car accident, the user of the autonomous refrigerator discussed earlier is unconscious. Lacking any new data regarding this condition, the fridge orders the food. Is it important that the user stays in "a state excluding conscious or free decision-making and expressing the intent", i.e. in a state excluding the possibility of making a valid declaration of intent? Or is it relevant that the AI which manages the fridge has concluded the agreement based on incomplete data?

If the AI were a tool, it would not be possible to acknowledge a juridical act as valid when the user is in a state excluding any capacity for making the declaration of intent, except from the cases described in Sect. 4.4. However, when the AI is fully autonomous, this cannot be a reason for rejecting the juridical act as void.[21]

In addition, it is not a simple task to apply the concept of error (mistake) to this situation. When an AI is truly autonomous, the user, as a rule, does not exactly know what the AI would do: in such cases, the will of the AI complements, substitutes or

[20] In this context the rules of Article 14 of the United Nations Convention on the Use of Electronic Communication in International Contracts adopted on the 23 November 2005, entitled Error in Electronic Communication, are not very useful. Especially that the Explanatory note by the UNCITRAL secretariat on the United Nations Convention on the Use of Electronic Communications in International Contracts (pp. 69–70) says:

> 211. At present, the attribution of actions of automated message systems to a person or legal entity is based on the paradigm that an automated message system is capable of performing only within the technical structures of its preset programming. However, at least in theory it is conceivable that future generations of automated information systems may be created with the ability to act autonomously and not just automatically. That is, through developments in artificial intelligence, a computer may be able to learn through experience, modify the instructions in its own programs and even devise new instructions.
>
> 212. Already during the preparation of the Model Law on Electronic Commerce, UNCITRAL had taken the view that that, while the expression "electronic agent" had been used for purposes of convenience, the analogy between an automated message system and a sales agent was not appropriate. General principles of agency law (for example, principles involving limitation of liability as a result of the faulty behaviour of the agent) could not be used in connection with the operation of such systems. UNCITRAL also considered that, as a general principle, the person (whether a natural person or a legal entity) on whose behalf a computer was programmed should ultimately be responsible for any message generated by the machine (see A/CN.9/484, paras. 106 and 107).
>
> 213. Article 12 of the Electronic Communications Convention is an enabling provision and should not be misinterpreted as allowing for an automated message system or a computer to be made the subject of rights and obligations. Electronic communications that are generated automatically by message systems or computers without direct human intervention should be regarded as "originating" from the legal entity on behalf of which the message system or computer is operated. Questions relevant to agency that might arise in that context are to be settled under rules outside the Convention.

[21] In our view, Poncibò (2022, p. 208) is wrong in stating that "it is clear that the lack of human psychology in AI contracting makes it difficult to apply the traditional remedies for defects in consent, such as mistake, fraud, threats and unfair exploitation".

even changes the will of the user. It is the risk borne by the user that sometimes the decisions of the artificial agent may differ from the decisions the user would make himself; this could be seen as a price the user has to pay for the comfort of having an AI. Exactly the same discrepancy would be observed if the agent of the user were human, so the user employs another human to perform duties in his stead, for example a housekeeper. Therefore, if the refrigerator buys a bottle of wine for Friday night and the user goes out because at the last moment he has been invited by friends, it does not mean that there was any mistake on the side of the AI or the user, or that the transaction should be invalidated. The only effect will be that the AI will not buy any wine the following week, because the bottle will stay in the fridge. Besides, legal transactions are usually based on the confidence in the declarations made by the parties. When an AI activated by the user makes a declaration of intent, it would undermine the safety and stability of future transactions to acknowledge the right to withdraw this declaration in the case of a discrepancy with the intent of the user. In practice, this would be equal to all declarations of intent made by the AI then being confirmed by the user if they are to be efficient.

In addition, situations can arise when the other party, being a human or a weaker AI, claims withdrawal from a legal transaction participated in by an AI because of supposedly vitiated consent or intention. This may be a result of the obvious cognitive and, in the case of humans, emotional inequality between parties. If our autonomous refrigerator looks for the best offers for its user, it may find ones which, when accepted, could be qualified as unfair exploitation or significant bargaining power imbalance which undermines the freedom of contract, and the AI may be charged with taking advantage of this state of affairs. It would be difficult to attribute the awareness of these circumstances to the user when existing standards and rules are to be kept. The same problem would be apparent when good or bad faith is at stake. Attributing bad faith or fault to the user when the AI acts autonomously, and whose actions are not ex definition predictable, would be possible only if some new mechanisms of attributability were introduced.

Once again, it appears that the "tool model" of AI, i.e. one which ignores the carrier and focuses on the parties of agreement, is not sufficient to account for any defects of declaration of intent when an AI decides by itself. In contrast, presenting an AI as an agent (representative) solves these problems and allows the AI to be included in the theoretical construction of the defects of declaration of intent. As Lewaszkiewicz-Petrykowska (1983), p. 18 notes:

> When the juridical act is concluded through the representative only the declaration of intent of this representative is an indispensable and required element of this juridical act. His declaration, and not the principal's attitude, is decisive for the existence of the juridical act. This declaration is also the subject of interpretation made for establishing the content and the correctness of the performed act. Therefore, the circumstances which could justify the undermining of the legal efficiency of the made declaration should be looked for in the declaration of intent of the agent.

Legal provisions usually explicitly describe the situation when the source of the defectiveness is the third person. And so DCFR does:

II. – 7:208: Third persons

(1) Where a third person for whose acts a party is responsible or who with a party's assent is involved in the making of a contract:
   (a) causes a mistake, or knows of or could reasonably be expected to know of a mistake; or
   (b) is guilty of fraud, coercion, threats or unfair exploitation, remedies under this Section are available as if the behaviour or knowledge had been that of the party.

(2) Where a third person for whose acts a party is not responsible and who does not have the party's assent to be involved in the making of a contract is guilty of fraud, coercion, threats or unfair exploitation, remedies under this Section are available if the party knew or could reasonably be expected to have known of the relevant facts, or at the time of avoidance has not acted in reliance on the contract.

Therefore, when an autonomous AI is an agent, any potential bases for the defectiveness of the declaration of intent must be sought in the actions of the AI system and not in the circumstances concerning the user. In such a case, when the AI's action is not defective and is performed on the basis of the properly-established mandate, even if it is not in line with the user's will, this action cannot be acknowledged as a defect of the declaration of intent according to the most commonly-binding rules.

Sometimes, some specific problems may arise when a human makes a declaration of intent to an AI system and later claims that the declaration was defective as result of the fact that the entity who acted on the other side, i.e. to whom the declaration was made, was an AI. These problems may be associated with:

(1) the identity of the party and the third person in the context of defects of will (in the future, an AI may play three roles: *viz.* of the party, of an agent, of the body of the legal person);
(2) a mistake as to the identity of the entity who acted, i.e. that an AI acted instead of a human being or vice versa;
(3) what to do about the unconsciousness of an AI in cases when the legal rules require conscious action for acknowledging a defect of intent.

5.5.2 Mistake and Fraud

Rules on the avoidance of a contract can be found in many civil codes and soft law instruments in contract law. They are usually constructed according to the following scheme mirrored by the DCFR:

II. – 7:201: Mistake

(1) A party may avoid a contract for mistake of fact or law existing when the contract was concluded if:
  (a) the party, but for the mistake, would not have concluded the contract or would have done so only on fundamentally different terms and the other party knew or could reasonably be expected to have known this; and
  (b) the other party;
    (i). caused the mistake;
    (ii). caused the contract to be concluded in mistake by leaving the mistaken party in error, contrary to good faith and fair dealing, when the other party knew or could reasonably be expected to have known of the mistake;
    (iii). caused the contract to be concluded in mistake by failing to comply with a pre-contractual information duty or a duty to make available a means of correcting input errors; or
    (iv). made the same mistake.

(2) However a party may not avoid the contract for mistake if:
  (a) the mistake was inexcusable in the circumstances; or
  (b) the risk of the mistake was assumed, or in the circumstances should be borne, by that party.

II. – 7:205: Fraud

(1) A party may avoid a contract when the other party has induced the conclusion of the contract by fraudulent misrepresentation, whether by words or conduct, or fraudulent non-disclosure of any information which good faith and fair dealing, or any pre-contractual information duty, required that party to disclose.

(2) A misrepresentation is fraudulent if it is made with knowledge or belief that the representation is false and is intended to induce the recipient to make a mistake. A non-disclosure is fraudulent if it is intended to induce the person from whom the information is withheld to make a mistake.

(3) In determining whether good faith and fair dealing required a party to disclose particular information, regard should be had to all the circumstances, including:
  (a) whether the party had special expertise;
  (b) the cost to the party of acquiring the relevant information;
  (c) whether the other party could reasonably acquire the information by other means; and
  (d) the apparent importance of the information to the other party.

In this context, it seems obvious that a human, or a legal person represented by a human, cannot plead a mistake which would be defined as a discrepancy between his own idea of the declaration he would have made, or would have wanted to make, and the actual declaration made by the AI. Therefore, the declaration of the AI should be qualified as if it is a declaration by the very party. As such, any potential discrepancy between the intent of the party and the declaration of the AI could only serve as a

basis for efficient avoidance of agreement if the general prerequisites of the mistake are realized.

When establishing whether these prerequisites were realized, a key role would be played by the identity and the characteristics of the AI recorded in the register. If the AI acted outside its registered aim (intended purpose), two legal solutions are possible: either AI may be acknowledged as not having legal capacity or the fact that the register is easily accessible may be a reason for recognizing that the second party could easily discover the mistake. For instance, our autonomous refrigerator orders two tons of butter, despite being registered as being suitable for the domestic use only. Such an order, and the agreement resulting from it, are obviously burdened with the defect. However, if we acknowledge that the source of this defect is a lack of legal capacity by the AI, better protection would be awarded to the person on whose behalf the AI acted, in our example, the user of the autonomous refrigerator. However, when the defect is qualified as a mistake, the risk is distributed between both parties: the user can claim the mistake if the other party knew or could reasonably be expected to have known of the mistake. When legal transactions are of a mass character, where the parties are in fact anonymous, choosing the first or second of these conceptions would give extremely different conclusions. However, assuming that there is a need to protect human autonomy, and that a human should not be involved in legal relations which are created by AIs and exceed human anticipations, there is a need for stronger protection than that given by the institution of mistake. If an AI acts in a different way than recorded in the register and does not realize the registered aim (intended purpose), its action should be recognized as being legally invalid as a whole. The damages arising as a consequence of these actions should burden the provider (producer) of this AI and not the user. The liability of the user cannot be a strict liability.

However, a different situation arises when AI acts without authorization. In this case, the institution of *falsus procurator* could be applied, as described earlier. Therefore, the rules on the mistake can be used only when AI acts according to its registered aim (intended purpose) and within the scope of authorization, e.g., a domestic refrigerator buys 200 g of cheese, but it makes an order which would not be made by the user himself: for example, it buys a cheese which is not liked by a user or to which the user is allergic. The withdrawal of such a juridical act would be possible according to general rules.

It should be noticed here that two kinds of situations are possible: first, a mistake made by an AI and second, a mistake made by a human using the AI. The second situation is not specific for problems concerning AI systems, and its qualification is not difficult. If a human using an AI, as a natural person or as a body of a legal person, makes a mistake, and the content of the declaration of intent made by the AI is the consequence of this mistake, the rules on the mistake should be applied. The fact that a stage exists "between" the will of the human and the declaration of intent where the will was completed, realized or clarified by an AI is not relevant. For instance, someone asks an electronic home assistant to find and buy an attractive trip for a wedding anniversary; the AI realizes the order and concludes the agreement, the exact content of which is unknown to the man. In the meantime, however, the spouse

of the user buys a trip for the same time. In such circumstances, a potential case for withdrawal based on the mistake should be evaluated, putting aside the fact that AI participated in the process of concluding the juridical act. In the same way, the fraud should be treated.

A different situation occurs when the mistake is made by an AI and not by the user. Let's imagine that based on the earlier behaviour of the user, the AI interpreted the command of the user incorrectly and ordered a trip which the user did not want to buy. This mistake could be a result of incomplete data or systemic reasons: the AI system is not perfect and is not able to predict the user's preferences properly. Such a situation is further complicated by the subjective aspect of the mistake; this subjectivity is indispensably present when the entity making the declaration of intent is a human, who is conscious, but the mistake must be objectivized when the entity is an AI, who is not conscious. When AI has all the data it needs for a correct declaration, it should be acknowledged that there is no mistake, even if these data are processed badly due to the imperfection of the system. However, the mistake may also embody an error *sensu stricto*, when the very act of declaration was disrupted. For instance, it is possible that an autonomous refrigerator may send the same order twice due to a problem with the internet connexion, or because of the failure of the system orders 100 kg of fruit instead of 1 kg, which is needed according to its calculation. Such an error should be treated as a mistake which allows the contract to be voided according to general rules.

It is also possible that a mistake by the AI system may be induced intentionally. Of course, while the methods would differ, the legal consequences should be analogical to when a human is defrauded. Very often, such an intentional mistake would consist of giving false or incorrect data; a simple example would be the case when AI buys a gold ring (according to the description included in the second party's offer) and it turns out to be copper.

A separate issue would be the possibility of avoiding a contract based on mistake by a human who is a party of the contract when the second party is an AI. Such a situation may entail new circumstances in addition to the known ones. A fundamental problem obviously concerns the mistaken identity of the contractor: when a human party does not know that he has entered into a relationship with an AI acting on behalf of some other person. This should be systemically excluded by imposing the informational burden on the AI. However, it cannot be excluded entirely that the person may have himself made a mistake. However, if the person was correctly informed that he is communicating with an AI, and that the AI acts within the scope of its legal capacity and authorization, the flaw of the person will not be qualified as a legally relevant mistake. In such circumstances, other protective legal instruments may be used, such as the right of withdrawal within some period of time after concluding the agreement.

When the human party is not correctly informed about the fact that the second party or his agent is an AI, the mistake arising as a consequence of this fact should be always qualified as a mistake induced insidiously (fraud), and that the lack of information should be acknowledged as important. Hence, the misinformed party should be allowed to avoid a contract based on these circumstances.

5.5.3 Threats

According to DCFR:

> **II. – 7:206: Coercion or threats**
> (1) A party may avoid a contract when the other party has induced the conclusion of the contract by coercion or by the threat of an imminent and serious harm which it is wrongful to inflict, or wrongful to use as a means to obtain the conclusion of the contract.

It is not doubtful that AI can formulate communications which may be qualified as a threat, in a traditionally-understood sense. Exactly like a human, an AI may say that it would cause harm to the second party if the contract were not concluded. However, while there is no way to prevent such situations in the case of a human, it should be possible to exclude such potential for threats at the construction stage. As such, it should be assumed that no AI capable of such actions which may be qualified as threat would be permitted to act and be registered. Furthermore, as noted above, when an AI acts without registration, or contrary to its registered intended purpose or scope, or if it acts according to the limits recorded in the register but when the information given in the registration process was false, such action should be recognized as legally not effective, i.e. as performed without legal capacity.

Of course, registration itself does not guarantee that the AI would not formulate a threat. Analogically to the situation of mistake, it may happen that a legally acting AI would act in such a way that, in a given set of circumstances, may be qualified as coercion or threat; in addition, it should be admitted that although AI is not conscious, it may nevertheless use coercion or threats. Therefore, the objective concept of coercion and threat should be accepted, where the aim of the action is reconstructed from the external behaviour itself.

A question may be put whether AI is capable of causing serious harm to a human, and whether its threats should be treated as a possible basis for voiding a contract. While this may depend on the concrete circumstances of a transaction, an AI may employ a wide range of possible strategies that may be interpreted as a threat. For instance, it may warn that if the second party does not conclude a contract with certain content, the AI would break the safety barriers in the party's accounting system and will change the records in a way which would cause problems with taxation for the party.

Much more interesting questions seem to be whether it is possible that an AI could be threatened by a human second party, and the AI makes the declaration of intent under the influence of such a threat, and whether in such circumstances, the user of the AI may void the contract. While such a situation is difficult to imagine today, it cannot be excluded in the future. In theory, even weak but advanced AI systems may be capable of evaluating the weight of such threats, for example, that the second party will cause the breakdown of an AI system, and the probability of its realization, and may change its choice to avoid such breakdown.

On the other hand, a specific kind of coercion would be where a human interferes with the AI's action, for example by modifying its code (hacking), to obtain the expected declaration of intent. Alternatively, the desired effect on the AI system may also be achieved by other technological factors, such as influencing magnetic field strength or electric current intensity. Such situations should also be qualified as coercion, which would entitle the user of the AI to void the agreement.

5.5.4 Unfair Exploitation

Especially interesting problems arise when examining the issue of unfair exploitation in contracts concluded with the participation of AI. Many legal systems include mechanisms which allow the contract to be modified or invalidated when a party contrives to conclude an agreement excessively beneficial for himself by intentionally making use of a specific circumstance concerning the other party. The model regulation of this kind is included in DCFR:

II. 7:207: Unfair exploitation

(1) A party may avoid a contract if, at the time of the conclusion of the contract:
   (a) the party was dependent on or had a relationship of trust with the other party, was in economic distress or had urgent needs, was improvident, ignorant, inexperienced or lacking in bargaining skill; and
   (b) the other party knew or could reasonably be expected to have known this and, given the circumstances and purpose of the contract, exploited the first party's situation by taking an excessive benefit or grossly unfair advantage.

Since the main element of unfair exploitation is intentionally (consciously) profiting from a certain condition of the other party, this again presents civil law with the problem of subjective circumstances. To qualify a situation as unfair exploitation, it is necessary that "the other party knew or could reasonably be expected to have known" about the condition. When an AI participates in legal relations, and in which it may be "the other party", the question arises regarding its "knowing" or "knowledge". In this matter, it is possible to take one of the three positions:

1. Firstly, because of the literal interpretation of the provisions, it may be assumed that the knowledge available to the contract's party, which is relevant for unfair exploitation, is exactly the knowledge available to the entity on whose behalf the AI acts. However, we are of the opinion that such a position, based on the "tool" aspect of AI, is not enough: when an AI concludes the contract, it makes the declaration of intent by itself, and the party of the contract, and is bound by it, may know very little, or even nothing, about the circumstances in which the concrete contract was concluded. Of course, such knowledge may be accessible by examining the data recorded by the AI, but it is counterfactual to assume that

this knowledge remains part of the mental content of the party who advantageously used the situation of the other party. Admitting such a position may cause that the institution of unfair exploitation would be impossible to apply.

2. A second theory exists based on a different position. Although it accepts the broader interpretation of the text of provisions, it also assumes that the subjective, volitionary elements are essential components of unfair exploitation. As such, it may be deemed that, theoretically, the knowledge, which is relevant to the institution of unfair exploitation, is available to the AI itself; furthermore, it can also be said that the action of an AI directed to exploit the other party's situation would also be relevant for the institution of unfair exploitation. But in fact, as long as AI is weak and lacks consciousness, it is impossible to attribute such behaviour to AI. In other words, the AI does not think, does not know anything, cannot act intentionally, and cannot exploit the other party's situation. According to this position, a charge of unfair exploitation requires some subjective element which cannot be attributed to the AI, for example, the mental attitude of the entity who violated some ethical standards. Such a position is correct in the sense that it precludes the analysis of AI as a moral subject. Since if it were assumed that AI has some knowledge and were able to use it intentionally for objectionable purposes, it would mean that AI was subject to some ethical demands it was capable of observing (*a contrario* from *impossibilum nulla est obligatio*). Besides, if AI had such capacity it would also be capable in a much wider context.
3. Finally, it is possible to take a third position which, in our opinion, is the most adequate. This position is also in some way universal because it would be suitable either when AI acts in its own name as a e-person, or when it acts as an agent of another entity within the scope of some authorization. According to this position, it is possible that an AI may engage in unfair exploitation or facilitate it, despite it not being conscious, because it is possible to give some objective standards of action. So, although a "normal" AI acts within the limits of its registration and the acknowledged legal capacity, it may exceed the acceptable way of acting in legal transactions. This takes place when an AI, while operating with the aim to conclude a contract, uses selected parts of accessible data concerning the specific position of the second party to give it or its principal an excessive benefit or grossly unfair advantage, and which would not be possible if the AI did not have this information. Such information could be that the party was dependent on or had a relationship of trust with AI or its principal, was in economic distress or had urgent needs, or was improvident, ignorant, inexperienced or lacking in bargaining skills. Assuming that the AI is capable of purposeful action, which is reasonable considering the current state of technical development, it should be admitted that the AI is capable of having and using accessible information resources. So, when its embedded aim (intended purpose) is to conclude maximally-beneficial contracts in its own or in someone else's name, arguably a legitimate aim in a market economy, it is possible that an AI may be acknowledged with unfair exploitation in specific circumstances. Indeed, excessive benefit or grossly fair advantage are, by definition, included in the notion of maximally-beneficial contracts.

However, three cases are possible here: (a) identifying opportunities for unjust exploitation may be encoded in the AI system by the constructor, or any other person; (b) identifying such opportunities may not be the preferred way of action embedded in the AI system, but there are no securities against it; (c) unfair exploitation is limited with the help of special tools embedded in the system forcing the AI to ignore any data on the specific situation of the second party. Possibility (c) may be made one of the criteria of AI registration, so when AI acts against them explicitly, as possibility (a), or implicitly, as possibility (b), its action should be recognized as illegal and culpable, within the concept of culpability proposed in Sect. 11.6.

Of course, preparing the technical securities (tools) mentioned above is not easy because of the vague and non-scalar character of such notions like "excessive benefit" or "grossly unfair advantage", but also "dependent", "relation of trust", "economic distress", "urgent needs", "improvident", "ignorant", "inexperienced" and "lacking in bargaining skill", especially when the second party is a professional participant of the market where the requirements of knowledge and skills are much higher than those expected from consumers.

Certainly, in the event of a concrete case of unfair exploitation by an AI being recognised, the consequences should be attributed to the entity in whose name or interest the AI acted (natural person, legal person or e-person if such is accepted in the given legal system). When it happens, there is no need to identify any subjective elements in the behaviour of that person, e.g. culpability, consciousness or unfair exploitation. The only condition is that the AI acted within the scope of its authorization. When the AI was authorized to conclude sales contracts concerning certain kinds of goods and concluded the sales contract on these goods under conditions of unfair exploitation, even if the authorization included the condition of the AI acting within the limits of the law, it cannot be acknowledged that AI exceeded the scope of its authorization; as such, the person on whose behalf it acted can avoid the consequences of unfair exploitation. So, where the appropriate legal provisions state "the other party" or "exploiter", this should be understood broadly as embracing the agent too, even if this agent is an AI; indeed, this is a binding principle in many legal systems today.

Therefore, assuming our autonomous refrigerator identifies an extremely good offer on food, which it is authorized to buy, if it has information that the reason for the sale is the economic distress of the seller, but nevertheless decides to buy the food, its action may be qualified as unjust exploitation. When the seller wants to avoid a contract, he has the right to do so, even if the user of the refrigerator did not know about the situation.

References

Books and Articles

Avila Negri SMC (2021) Robot as legal person: electronic personhood in robotics and artificial intelligence. Front Robot AI 8:article 789327. https://doi.org/10.3389/frobt.2021.789327

Beran K (2020) The concept of juristic person. Wolters Kluwer, Prague Warsaw Bratislava Budapest

Bort J (2017) How Salesforce CEO Mark Benioff uses artificial intelligence to end internal politics at meetings. 19.05.2017 Business Insider. https://www.businessinsider.com/benioff-uses-ai-to-end-politics-at-staff-meetings-2017-5?IR=T, last access on the 4th of August 2022

Burridge N (2017) Artificial intelligence gets a seat in the boardroom; Honk Kong venture capitalist sees AI running Asian companies within 5 year. Nikkei Asia Review. 10.05.2017. https://asia.nikkei.com/Business/Artificial-intelligence-gets-a-seat-in-the-boardroom, last access on the 4th of August 2022

Chopra S, White LF (2011) A legal theory for autonomous artificial agents. The University of Michigan Press, Michigan

Dahiyat EAR (2020) Law and software agents: are they "Agents" by the way? Artif Intell Law. https://doi.org/10.1007/s10506-020-09265-1

del Val Talens P (2017) Corporate directors: in search of a European normative model for legal persons as board members. Eur Company Financ Law Rev 14(4):609–636. https://doi.org/10.1515/ecfr-2017-0028

Drukarch H, Fosch-Villaronga E (2022) The role and legal implications of autonomy in AI-driven boardrooms. In: Custers B, Fosch-Villaronga E (eds) Law and artificial intelligence. Regulating AI and applying AI in legal practice. Asser Press-Springer, Cham, pp 345–363

Filatova N (2020) Smart contracts from the contract law perspective: outlining new regulative strategies. Int J Law Inf Technol 28(2020):2017–2242. https://doi.org/10.1093/ijlit/eaaa015

Frier BW, McGinn TAJ (2004) A casebook on roman family law. Oxford University Press, Oxford

Habibzadeh T (2016) Analysing legal status of electronic agents in contracting through interactive websites: comparative study of American, English and EU laws developing Iranian legal system. Inf Commun Technol Law 25(2):150–172. https://doi.org/10.1080/13600834.2016.1186361

Katz A (2010) Intelligent agents and internet commerce in ancient Rome, society for computers and law. https://www.scl.org/articles/1095-intelligent-agents-and-internet-commerce-in-ancientrome, last access on the 4th of August 2022

Kerr I (1999) Spirits in a material world: intelligent agents as intermediaries in electronic commerce. Dalhousie Law J 2

Kerr I (2001) Ensuring the success of contract formation in agent-mediated electronic commerce. Electr Commer Res J 1:183–202

Lewaszkiewicz-Petrykowska B (1983) Problem wad oświadczenia woli w czynności prawnej dokonanej przez przedstawiciela. In: Rembieliński A (ed) Studia z prawa cywilnego. Księga pamiątkowa dla uczczenia 50-lecia pracy naukowej prof. dr hab. Adama Szpunara. Warszawa – Łódź

Linarelli J (2022) A philosophy of contract law for artificial intelligence: shared intentionality. In: Ebers M, Pancibò C, Zou M (eds) Contracting and contract law in the age of artificial intelligence. Hart Publishing, Oxford, pp 59–80

Mucha-Kujawa J (2017) Teoretycznoprawne aspekty przedstawicielstwa organizacyjnego jako pragmatycznego sposobu reprezentacji osoby prawnej. Studia prawnicze PAN 3(211):2017

Pagallo U (2010) Robotrust and legal responsibility. Knowl Technol Policy 23:367–379
Poncibò C (2022) Remedies for artificial intelligence. In: Ebers M, Pancibò C, Zou M (eds) Contracting and contract law in the age of artificial intelligence. Hart Publishing, Oxford
Wein L (1992) The responsibility of intelligent artifacts: towards an automation Jurisprudence. Harv J Law Technol 6(1992)
Zou M (2022) When AI meets smart contracts: the regulation of hyber-autonomous contracting systems? In: Ebers M, Pancibò C, Zou M (eds) Contracting and contract law in the age of artificial intelligence. Hart Publishing, Oxford

Documents

Council Regulation (EC) No 2157/2001 of 8 October 2001 on the Statute for a European company (SE) with amendments, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32001R2157, last access on the 4th of August 2022
Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (recast) L 173/349 (MIFID II). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0065, last access on the 4th of August 2022
United Nations Convention on the Use of Electronic Communications in International Contracts (New York, 2005) https://uncitral.un.org/sites/uncitral.un.org/files/media-documents/uncitral/en/06-57452_ebook.pdf, last access on the 4th of August 2022

Chapter 6
Personal Interests of AI

6.1 Introduction

For the sake of clarity, we should differentiate between *personal interests*, as a concept of civil law, and human rights, personal rights and fundamental rights, which are used more in a broader sense as concepts of public international law, European law, constitutional law and legal philosophy. Personal interests are primarily the attribute of an individual, while human rights, fundamental rights or personal rights are only attributed to an individual. The former is applied in legal relations of a horizontal character, i.e. those between individuals, while the latter are engaged in those of a vertical character, i.e. between the individual and the state. However, these categories overlap and are interconnected. Personal interests are often justified based on the fundamental rights of a man, mainly his dignity. Both personal interests and human rights are intertwined with the physical and mental integrity of an individual.

The command to respect fundamental rights is acknowledged as the foundation of human rights protection and is, at a fundamental level, common to all democratic legal systems. The Charter of Fundamental Rights of the European Union proclaims *inter alia* that human dignity is inviolable (Article 1) and everyone has the right to life (Article 2); it also enshrines physical or mental integrity (Article 3), liberty (Article 6), respect for private and family life, home and communications (Article 7), protection of personal data concerning an individual (Article 8), the right to freedom of thought, conscience and religion (Article 10), freedom of expression (Article 11), freedom of peaceful assembly and association at all levels (Article 12), the right to education and access to vocational and ongoing training (Article 14), the freedom to choose an occupation and to engage in work (Article 15) and the right to property (Article 17). It also includes the freedom to conduct a business (Article 16). In many states, sets of rights are included in roughly similar forms in their constitutions or equivalents, while personal interests tend to be included in the acts serving as the

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_6

basic source of civil law, mainly in civil codes. This kind of model regulation is also included in the DCFR which, in the chapter "Principles", page 81, declares *inter alia*:

> 32. Protection of the person. A particular concern of non-contractual liability law is the protection of the person. The individual stands at the focus of the legal system. A person's rights to physical wellbeing (health, physical integrity, freedom) are of fundamental importance, as are other personal rights, in particular that of dignity and with it protection against discrimination and exposure. Injuries to the person give rise to non-economic loss besides economic loss; the former also deserves compensation.
>
> 33. Protection of human rights. The non-contractual liability law of the DCFR has the function primarily (albeit not exclusively) of providing "horizontal" protection of human rights – that is to say, a protection not vis-à-vis the state, but in the relation to fellow citizens and others subjects to private law. This protection is provided in the first instance by the claim to reparation for loss suffered, but is not confined to that. [. . .]

The important consequence of the described relationship between rights, which are strictly human, and personal interests, which can also belong to some non-human entities, is that the latter constitute a general category which includes attributes belonging to a concrete individual; these cannot be transmitted to another individual or be renounced, either by the individual or by anyone else. If this is so, they are essentially the attributes of objects specified as to their identity, i.e. who have their own identity: personal interests, by definition, cannot apply to objects specified as to their kind, i.e. to an entity who only exists as an anonymous element of a category. This may be the reason why courts have often refused to protect personal interests when the potential infringement referred directly to an entire category of persons.[1] It was also the reason why in the "monkey selfie" case, in which PETA filed a copyright lawsuit against D. Slater and Blurb Inc., the monkey portraited on the photos was endowed with a name (Naruto). Naming the monkey did not protect the plaintiff, however, against the argument that he sued on behalf of the wrong monkey: although the lawsuit described a six-year-old male crested macaque, Mr Slater, the photographer, in his book, described the monkey as female.[2] Another famous case confirming the necessity of identity concerns the anonymous street artist Banksy, whose trademark was cancelled because of his anonymity. According to the European Union Intellectual Property Office,

> The fact that Banksy has chosen to be anonymous and cannot be identified would also hinder him from being able to protect any such copyrights accruing to his art. [. . .] he cannot be identified as the unquestionable owner of such works as his identity is hidden; it further cannot be established without question that the artist holds any copyrights to a graffiti. [. . .]

[1] E.g. in Poland the case of the press article containing the information that during the II World War the Polish engine drivers had collaborated with Nazis. The publisher was sued by the labor union representing the engine drivers for the infringement of the dignity, honour and good name of Polish engine drivers. The case was dismissed. (sentence of the Appellate Court in Warsaw of 29.09.2020 V Aca 823/19), https://sip.lex.pl/#/jurisprudence/522694791, last access on the 4th of August 2022.

[2] Kravets (2015).

> Therefore, the filing of a trade mark cannot be used to uphold these rights which may not exist, or at least may not exist for the person claiming to own them (Decision on Cancellation No 33843 C).

On the contrary, in a case decided in the Netherlands concerning compensation for the cost of medical treatment of an injured cat, the judge justified the high amount of the damages as follows:

> [. . .] companion animals such as dogs and cats are regarded a part of the family within which they fulfil a certain affective/emotional role. In view thereof, not every cat is the same [. . .].[3]

These disputes, sometimes incomprehensible for the audience, demonstrate the importance of the issue of identity. This issue becomes heightened when reflecting on the personal interests of AI, a topic which will be discussed further in this chapter. During the discussion, it will also become clear that these personal interests, which deriving from fundamental or human rights, are not applicable to AI, at least in the sense which is used in the hitherto acquis; it will also be emphasized that this state of affairs is justified on two bases: safety and non-adequacy.

A quite separate issue is that the existence of personal interests should be associated with a being endowed with some scope of legal personality. As such, it is accepted that these interests may be attributed to juridical persons, at least in some form, despite them not being human; however, in Western legal systems, personal interests of a human are lost at the moment of death, as they are considered to die together with personhood. Interestingly, while some values connected to former persons, i.e. those who are dead but previously alive, may migrate to other living persons, for instance the family members of the dead, this is not the case for personal interests.

So, if an entity has no personhood, it cannot have personal interests.

It was explained in Chap. 2 that legal subjectivity or legal personhood can differ between different kinds of entities, and that this variation is determined by their capabilities to participate in social life, with the emphasis on *human social life*, and by their social value. The legal subjectivity of humans can coexist with other forms of subjectivity, such as that of juridical persons, which is similar to the human form but with a different scope. Legal subjectivity is sometimes attributed to some immaterial or biological entities, albeit with a specific scope, and may also be attributed to animals and AI in a prospective sense. Therefore, it is in fact a fallacy to speak of *subjectivity* in general; rather, it is better to think in terms of a series of *subjectivities*, the content of which can be identified only based on the entirety of the legal and real situation.

The violation of personal interests may result in damage to both property and, more seriously, non-material objects. In addition to health, freedom and reputation, all legal systems protect the moral rights to works, such as the right of authorship. Some such interests, such as health, are strictly specific to biological entities, but others are not. For this reason, further reflection is needed, not only on whether it is

[3] Bernet Kempers (2021), p. 61.

justified to attribute personal interests to AI, but also to consider what these personal interests should look like. Is it possible to form an analogy to human personal interests?

Unlike animals, AI is not genetically or structurally similar to humans, and as such, it does not demonstrate emotions and feelings that approximate those of humans. AI, or the so-called *weak AI*, as underlined in the Introduction, does not think in the way human thinks, has no emotional attitude to its own actions, has no consciousness and feelings, and does not reflect on its individual future or how to shape it. It can pursue aims but it cannot create these aims by itself. For now, the aims of an AI are imposed by its creator, which is, at least for now, a human; while future AIs and their aims may be created by another AI, ultimately, at the end of this chain, all would be created by humans. Is it therefore reasonable to accept that, even if AI is accepted to have some legal subjectivity, it can also have its own personal interests? In our opinion, the scope of any potential personal interests awarded to AI must, at least to some extent, arise as a consequence of the scope of personhood it would be given, or serve as a reflection of it. As both legal subjectivity and personal interest derive from social relations, they function in parallel and concurrently, exerting a mutual influence on each other. Furthermore, as such relationships are so complicated and intertwined, it would be pointless to ask which of the two is conceptually and temporally the precedent.

As it was mentioned before, if an entity has the status of a legal subject, it is so because it participates in social relations and has some intrinsic or instrumental social value. It should also be remembered that both values may vary qualitatively and quantitatively, and may be represented to varying degrees in an entity. Intrinsic value is highly cherished, and in the case of human beings, allows them to be identified as legal persons and to be recognized as having distinct identity. It is not by accident that the right to identity, which encompasses name, surname, date of birth, gender and nationality, exists as a human right.[4] The personal interest overlapping with this human right may be named differently; e.g. in Polish Civil Code Article 23 it is called "surname or pseudonym".

However, while the attribution of instrumental value to an entity does not force any action on the part of lawmakers, if this value is significant (qualitatively or quantitatively) it may influence the reasonable lawmaker to make some proper and adequate decision as to the legal personhood of the entity, its scope, and any indication symbols identifying it. For example, the tool for labelling the identity of a juridical persons is their name, which in Polish Civil Code Article 43^5 Section 1 is called "business name" ("firma") and is strictly determined in the following provisions. The justification of these decisions is strictly utilitarian, compatible with physical, social, and legal reality of the given country or countries.

It should be noted that some situations may appear when the candidate to legal personhood is of both simultaneously intrinsic and instrumental value. For example, while cooperative societies are very useful tools for organizing society, they are also

[4] de Varennes and Kuzborska (2015).

of intrinsic value as organizations whose aim is to facilitate mutual help and cooperation. As an indication of how strongly these values are cherished, it may be said that while the intrinsic value of cooperative societies is certainly lower on the scale than the intrinsic value of a human being, their utility value is quite high: it is typically assumed that an organized group of people can do more than one person alone.

All these preliminary deliberations indicate that the nature of the personal interests of AI indeed lies within its situational character: it is first necessary to identify the actual social relations in which a given AI participates and the actual value it has for them; it is hence also necessary to examine the properties of a certain kind of AI, and the interests of people towards it. Only on these bases is it justified to determine whether a place exists for the personal interests of a certain AI. It is not reasonable at all to apply the concept of human personal interests to AI directly, because human personal interests are based on human rights (fundamental rights), which lie within the domain of humans and humans alone.

A certain starting point may be the observation that, unlike other non-human legal persons, some AIs are capable of autonomy; as such, they can choose between various options to achieve an assigned goal and thus obtain a result which is not programmed, predicted or even predictable by a man. This capability imparts an individual character to a given AI, one that is separate from its creator or trainer, and could serve as a basis for recognizing its identity. Such an identity is immanently connected to the existence of this, and only this, AI and, when recognized by a legal system, becomes inherently its personal interest.[5]

However, another problem is whether this identity and its accepted symbol would be protected by the law, especially by civil law; administrative law may protect such a symbol for different reasons. Certainly, recognizing some personal interests of AI does not necessarily entail giving them some direct legal protection, such as the capacity to sue when its personal interest was infringed. A number of possible regulations exist in this case. AI is not conscious, and, unless specifically designed otherwise, it does not have to be aware of its legal or real interests, or its possible infringements, even on the level of pure knowledge. In such a case, giving AI the capacity to sue has no sense. However, this would not prevent other legal subjects from acting on their own behalf, or on behalf of the AI, from filing certain claims deriving from the personal interests of AI; alternatively, they could plead the AI's personal interest to refute the claims of others. The procedural argument of this kind is easy to imagine in practice: the defendant in a personal interest case could argue that although he infringed personal interest, it was that of the AI, not the plaintiff. In other words, the only obvious consequence of recognizing the personal interests of an AI, assuming this is made possible by the lawmaker, would be that they would

[5] Of course, the legal system must decide about the symbol indicating such AI. To make the system coherent and safe, symbols should be unique just like the identity numbers applied in many countries to identify individual people. The most probable solution would be a combination of letters and numbers. The symbol of an AI should be recorded in the register, which is the domain of administrative law and will not be considered in this book.

influence the scope of rights and duties of other persons. As an example, it is worth considering the right to authorship. While it is possible that an AI's right to authorship may be recognized by the legal system, i.e. confirming that an AI can be an author of a work and it should be identified as such, this does not imply that an AI could have a claim if its right to authorship were infringed. However, if a given AI were identified as the author of a given work, this would prevent a third person from effectively claiming to be the author; for instance, the creator or user of an AI system would not have such a claim. Restricting a domain to AI, even if this domain is not directly protected by the law, narrows the scope of rights held by others.

However, the legal protection of AI's personal interests, either all or some, is not excluded; also, it would not be impossible for an AI to sue on its own behalf. Indeed, in a legal and technological reality more developed than today, such a solution would be economically justified. If, at the beginning in the simplest cases, the legal system were technically and structurally prepared for some automatic and autonomous court procedures and for the participation of AI in such procedure, AI could sue and be sued without the active participation of humans, either with the permission of its owner or, in the event it were a separate legal person, without it. Even so, it would be advisable to restrict the scope of this kind of process to AI only, for the sake of security for humans. These issues, however, are not the subject of this monograph, and they will not be discussed any more.

It is important to note that by recognizing some personal interests for AIs, humans would lose some degree of freedom, and this effect should influence the whole conception of AI personal interests.[6] Exactly as in the case of legal subjectivity, what is described in Chap. 2 AI personal interests should be constructed in a granular ("pricking") fashion. In contrast to the Western dominant concept of human personal interests,[7] AI would be allowed to have only these personal interests which would be explicitly affirmed by a given legal system. There is no place for AI to have either general legal subjectivity or generally-defined legal personal interests. The personal interests of AI should be allowed on a pointwise basis, i.e. only in cases when they do not limit the fair rights of human beings.

[6] Similarly, today the idea that the protection of animals may limit the rights of humans is not unknown. In Germany since 2002, it has been possible for them to be weighed against constitutionally protected human interests. Bernet Kempers (2021), p. 44.

[7] In the Polish Civil Code it is expressed by the phrase "Personal interests of a human being, such as in particular. . ." (Article 23).

6.2 The Possible Types of Personal Interests of AI

6.2.1 Existence and Procreation

Safety dictates that humans need to retain complete control over non-human entities, machines and AI included; on this basis, no warranty of existence should be acknowledged for AI, at least for now, while the existence of a conscious AI lies more within the realm of science-fiction. We speak of *existence* and not *life*, because being a computer system, at least for now, AI is not *alive* in a biological sense. Furthermore, as an AI is based around software, the very existence of such system is not so important; what is more important is its possibility for action and its ability to realize its aims. Although the code stored on any medium or hidden in a desk drawer may exist, it does not influence anything. So, as a matter of fact, when discussing some potential interest of the existence of AI, it is necessary to consider not only its very existence but rather its existence in a form which allows it to realize its given aims, such as the AI installed in some hardware.

The non-biological nature of AI is an important consideration, as it is biological life that is cherished and protected by Western legal culture. While the degree of this protection depends on many factors, all forms of biological life, even the smallest, are regarded as having both intrinsic and utility value. Of course, human life is regarded most highly; however, this does not exclude that other living creatures may also have intrinsic value. For example, in the Netherlands, according to the Animal Protection Act and the Civil Code, animals are treated as having intrinsic value and even those regarded as property are distinguished from non-intrinsically-valued things. [8]

In contrast, in the case of other non-biological entities, such as juridical persons, their very existence is not protected, even if these entities have some legally-acknowledged personal interests. Various legal instruments exist that are used to maintain the existence of juridical persons,[9] but they are the consequence of having utility and not intrinsic value, and as such, they offer far less protection than personal interests. This difference between biological and non-biological (artificial) existence is fundamental for the issue examined in this chapter, as well as for many other issues touched on in this monograph. As it is artificial, and in the sense that it is non-biological, AI has no "right to life" or any claim to its own existence (executed

[8] Bernet Kempers (2021), p. 46.

[9] E.g. the § 29 BGB commands that: "To the extent that the board is lacking the necessary members, they are to be appointed, in urgent cases, for the period until the defect is corrected, on the application of a person concerned, by the local court [Amtsgericht] that keeps the register of associations for the district in which the association has its seat." The Polish Civil Code commands that "If a legal person cannot be represented or manage its own affairs due to the lack of organs or the incomplete composition of the organs authorized to represent it, the court shall appoint a curator for it. [. . .] The curator shall immediately undertake actions aimed at appointing or supplementing the composition of the organ authorized to represent, or if necessary, to liquidate, the legal person". (Article 42 section 1 and 3).

by itself or by humans acting on its behalf). No matter what values it is capable of creating, be they fantastic novels or pertinent medical diagnoses, and the potential importance of its social role, such as concluding contracts on behalf of mentally-disabled people, managing large city infrastructures or performing psychological care of policemen and firemen, the possibility of its action must wholly depend on decisions made by humans, or more precisely by those who are authorized to do so. Any decision to disable the AI or even destroy its code cannot be blocked in any way by the AI itself.

In the same way, it is not acceptable for an embodied AI, such as in a robot, to have any personal interest based on the right to existence, understood either directly or metaphorically. The very fact that a given AI is placed in some physical object (machine), metaphorically called a "body", does not afford it any "right" or claim or interest of this robot to keep its existence or to prevent its own destruction against the will of authorized humans.[10] It can be loosely said that, to some extent, Isaac Asimov was correct when he established the Second Law of Robotics, whose sense is that a robot must protect its own existence, as long as such protection does not conflict with the prohibition of injuring a human or letting him be injured, or conflict with obeying the commands of humans.[11] It should be remembered that, even if a robot had some personal interests, regardless of their content, their source and justification would not lie in the "body" but the "brain", i.e. the AI. This is the consequence of the fundamental assumption, accepted herein, that the legal situation of AI should regulated by private law, and the law in general, no matter whether it is embodied or not. Including AI in a body, human-shaped or not, does not justify the addition of any emergent or added value resulting in the body bearing more rights than the AI itself.[12] The situation would certainly change if AI were to be integrated into a biological body, human or animal; however, in our opinion, the technical, ethical and legal obstacles are too great for this to happen in the near future.

It is possible, though, that the situation may look different in other legal and cultural contexts:

1. In some cultures, certain objects which are not regarded as living or autonomous by Western culture (e.g. mountain or river) may be endowed with legal subjectivity and even some personal interests. This usually happens when these cultures maintain the belief that such objects are living or powerful. If so, it is possible that a similar argument could be made for AI or robots; unfortunately, no culture is currently known to maintain the belief that AI or robots are living and entirely autonomous. While this may change in the distant future, such ideas seem to be impossible today in European legal systems and *ius civile* and hence will not be considered here.

[10] However, in certain situations, e.g. when there is a strong anthropomorphisation of robots, people are inclined to grant a kind of 'right to life' to machines and even place them above the interests of humans, cf. Mamak (2021).

[11] Cf. Sect. 11.1 where the relevance of Asimov's Laws for the law is examined.

[12] Cf. Księżak and Wojtczak (2020).

2. In some circumstances, because of cultural, legal but also political or economic context, robots or a given robot may receive some legal status, interests or rights, as if they were humans. The first and best example of this kind of case is that of an android called Sophia which, as it was publicly declared, received the status of citizen of the United Arab Emirates. As Sophia clearly lacked any capabilities to use this status, its scope and associated rights or interests are established by arbitrary decisions of the state, which realizes some of its goals in this way: Sophia became a citizen of Emirates for promotional purposes. However, this seems to be superfluous and even pernicious, because, apart from the many dangers to the human individuals and society, indicated herein, such an approach tarnishes the existing legal tradition and reduces the coherence of the legal system. The case of Sophia has even been accused of devaluing human rights.[13]
3. The existence of some forms of AI, especially embodied ones, may be legally protected; however, such protection would not be owed to the AI itself, but rather as an element of the personal interest of a human being connected to an AI or robot with some kind of personal relationship. Such identification and execution of an analogical human personal interest, e.g. a person with an affectionate bond towards an animal, is possible today in some legal orders. Bernet Kempers (2021, p. 59) describes several court cases in France and Belgium where the owner of a killed animal was awarded compensation for moral damage; the judgment was supported by the argument of subjective and affective disadvantage of a human losing his pet, with its source in the emotional bond with the animal. It is hence not unimaginable that in some circumstances, a human may have the legal possibility to block a decision to disable an AI, e.g. a social robot, even if this decision is made by an authorized person, such as its owner, because of the justified emotional attachment to this AI, should the AI for example, be this person's caregiver. It should be noticed that such a possibility does not exceed the existing provisions of *ius civile*, at least in some legal systems, and depends above all on the scope of protection granted in a given legal system to the non-property interests of a human. In these legal orders, where such protection is broadly understood, and can include emotional attachment to animals or even non-living objects, including attachment to AI would not be unusual. Ergo, paradoxically, even in these spheres where relations between human and AI are the strongest, such as friendship or love felt towards robots or avatars, the personal interests of AI are not at stake. However, the increasing proliferation of such phenomena, as well as their growing intensity and social commonness, may trigger the personalization of AI and evoke new legal regulations.[14]
4. A problem *prima facie*, resembling the above but quite different, may arise when the special relationship of adjustment is generated between a human and an AI, and more rarely, animals such as a seeing-eye dog, but this is another case; in such cases, the given AI becomes individualized through its interaction with a concrete

[13] Hart (2018).

[14] Kaplan (2016), *passim*.

human and the exposure to his exclusive knowledge or special skills. Notwithstanding the subjective attachment of a human towards AI, objectively both an AI and its accompanying human shape each other. As a consequence, the given AI becomes the human's "own", regardless of whether it is officially his property, because it becomes perfectly adjusted to the needs of the user. This phenomenon may take different forms depending on the form and complexity of the system. If the specification of an AI encompasses special values which are to be maintained or developed by the AI, such as support of older people, psychological treatment or help for people with some mental disabilities, its existence may become an issue far exceeding simple property issues and entering the non-property legal sphere. For instance, resetting an AI which for some time helped a suicidal person may endanger that person's life. Such situation does not create any right for AI to further exist in an untouched form, but such a right may arise, at least in some situations, for the user. The need to protect some fundamental rights, such as the rights to life or health, may require significant modification of license agreements or other legal titles determining the rights of an AI.

5. It is possible that an AI's personal interests may arise as a consequence of the human personal interest to have an unimpeded relationship with AI; however, in other cases, it may be necessary to develop a conception of AI personal interests where a form of behaviour towards an AI or robot would be prohibited by some accepted social rules. In such cases, the purpose of any prohibition would not be to protect of AI or its user, but to protect the values held by society. For instance, some forms of violent, abusive or discriminative behaviour towards robots may be prohibited to reduce or prevent their occurrence against humans.[15] In this case, such prohibitions cannot either be considered as a manifestation or proof of AI's personal interests, such as the right not to be treated violently, abused or discriminated against. However, it is not excluded that, in certain circumstances, it may serve as a starting point to awarding AI such pointwise legal subjectivity.

If the personal interest of existence and acting ("life") is not recognized, the personal interest of reproduction (copying) should also not be recognized. It is crucial for the safety of humans to maintain the power to determine whether code can be copied and disseminated. The capability of an AI to reproduce, which from the technological point of view, may take place to an unlimited extent, should be controlled from the beginning, as well as in the process of registration. The certification and registration demands should determine whether, and in what circumstances, copying is allowed and how to identify the copies of the code. AI cannot have the right to exist, reproduce or control its copies (what may be metaphorically perceived as "having a family"). AI has no claim to "raise" its copies according to its ideas, if training of AI systems may be metaphorically seen as an analogon of a human raising children.

[15] Cf. Darling (2016), Ryland (2021), Danaher (2017), Navon (2021).

6.2.2 *Personal Interests Related to Consciousness, Emotions and Embodiment*

It is unimaginable nowadays to recognize any personal interests of AI which could parallel human personal interests which have their source in dignity. It is not controversial today that human dignity encompasses the idea that every human being is of intrinsic value, which should never be diminished, comprised or repressed by other humans or any other entities, or by new technologies like AI systems.[16] Moreover, many people are convinced that dignity is a fundamental value, not reducible to any other more elementary features, unalienable and sacred, which may be ascribed only to humans as moral persons having the autonomy and capability to make free choices, and pursue personal fulfilment. The consequence of such a position is strong resistance against even the smallest possibility of giving AI any kind or any scope of dignity now and in the future, regardless of its development.

However, in the last few decades, many new ideas, social and scientific movements have refuted the anthropocentrism of previous years and attempted to assign a greater priority to natural phenomena, such as the climate, the earth and its animals; such attempts have also been made in the legal field. For example, the Swiss Animal Welfare Act of 16 December 2005 explicitly protects not only the welfare but also the dignity of animals (Article 1), where dignity is defined as follows (Article 3 letter a):

> Dignity: Inherent worth of the animal that has to be respected when dealing with it. If any strain imposed on the animal cannot be justified by overriding interests, this constitutes a disregard for the animal's dignity. Strain is deemed to be present in particular if pain, suffering or harm is inflicted on the animal, if it is exposed to anxiety or humiliation, if there is major interference with its appearance or its abilities or if it is excessively instrumentalized.[17]

The intentional or negligent disregard of animal dignity is recognized in this act as a criminal offence (Article 26).

Therefore, it is possible that the human monopoly for dignity may be questioned. In fact, this is already the case in the field of Philosophy,[18] and as demonstrated by the piece of legislation cited above, even the legal sphere does not appear to be completely resistant to the possibility. But is the dignity of an animal, if it exists, identical to the dignity of a human being? And if dignity were one day to be attributed to an AI, would it be the same as that attributed to a human?

Firstly, dignity is a normative concept, and a legal normative concept when used in the field of the law, and as such, it has some prescriptive consequences of a legal character. As A. Etinson (2020), p. 365 notes:

---

[16] McCrudden (2008).

[17] Translation from https://www.globalanimallaw.org/downloads/database/national/switzerland/Tierschutzgesetz-2005-EN-2011.pdf, last access on the 4th of August 2022.

[18] Many philosophers object this so called "speciesism" Cf. Meyer (2001); Singer (2009), pp. 309–356.

> If we put abstract theory aside for a moment, and look instead at our concrete ("applied") judgments about what human dignity practically requires, and when it is violated or most at stake, we see that it is preoccupied not simply with moral status (or even specific moral goods like autonomy or inviolability) but with *social* status – with "honoring" a person, as opposed to humiliating or degrading them.

The author also reflects on the dictionary meaning of the word "degrade", given as "to reduce from a higher to a lower rank, to depose from a position of honour or estimation", and identifies three general ways of perpetrating, intentionally or not, this kind of harm: disrespectful attitudes, expressions of disrespect and the loss of status markers. He also asserts that, following Kolnai (1976) and Rosen (2012), human dignity is only one species of dignity, and is simply different to others, even those belonging to human beings, e.g. parental dignity, academic dignity, judicial dignity and so on. He goes on to propose that human dignity is connected to a human's basic status as an equal in society, or as a fellow human being, whatever position this human holds.[19] If so, it should be noted that the meaning and the demands of dignity are generally dependent on the domain and circumstances in which it functions, as well as the also time and place. For example, the acceptable standards of living were much lower just after the Second World War than today, and as such, the standard of dignity was much lower; in addition, when we speak of the dignity of animals, as in the Swiss Animals Welfare Act, its form is appropriate to a given species. While no one would regard being kept and bathing in a manure as humiliating treatment for a dung beetle, this situation would not be acceptable for a human. So, reflecting on the problem of the probable dignity of AI, it is important to consider that it cannot exist as a calque of human dignity or animal dignity, but as a form of dignity suitable for an AI in a given time. Of course, its precise nature is difficult to determine, and requires another important point to be considered.

Even if the human monopoly for dignity is being disrupted, a limit to this disruption exists. Firstly, any serious discussion regarding such changes concerns not only living creatures, but also vertebrates: these are the focus of the Swiss act. Why vertebrates? Because all vertebrates are in a lower or higher degree conscious; all of them are probably sensory-phenomenally and affectively conscious; they may have a primary self (all vertebrates), a core self (reptiles, birds, mammals, homo sapiens), self-consciousness (birds, mammals, homo sapiens) and a narrative self (homo sapiens).[20] Similarly, if the potential for dignity among AIs is to be taken seriously, an AI should at least be conscious to some degree. Until this happens, and today it seems more the preserve of science-fiction than scientific perspective, it is even difficult to say what this form of dignity would consist of. It could not be linked with the capability of an AI to make autonomous choices, as these choices are limited by the values imposed by humans.

Thus, as long as the paradigm of civil law, and the law in general, attributes dignity to biologically living and conscious creatures, with human beings in first

[19] Etinson (2020), p. 372.

[20] Fabbro et al. (2015).

place, AI cannot be endowed with personal interests arising as the direct consequence of dignity, i.e. the personal interests connected with consciousness, or the psychological or physical structure of an entity. While AI does not know that it exists, there is no sense in protecting its existence; likewise while it does not feel pain, sadness or love, its embodiment, or relations with other entities, there is no sense in protecting it against phenomenally unpleasant experiences. These reasons alone would be sufficient to exclude the possibility of acknowledging such personal interests as health (mental or physical), personal immunity and integrity, or emotional bonds with another AI or a human, whatever their equivalent name may be.

Hence, the personal interests of an AI cannot include the integrity of its code, not only to ensure the safety of humans, but also the features of the AI itself rule this out. The power to change its own code cannot be awarded to an AI itself, either for corrections or for actualisation. Any damage to the code, or its distortion, retardation or limitation, even if done illegally, cannot be the source of any claims based on personal interests. Even if AI were acknowledged to be a legal subject capable of filing a suit, the source of its claims may only be property damage of real economic value. And if AI were the property of some natural or legal person, the claims are attributed to them. The same rules concern embodied AIs: they are robots, both within their code ("brain") and material medium ("body").

The presented position would not be affected by the increasingly common phenomenon of humans forming emotional bonds with AIs. This is particularly common for special types of AI, or more likely robots, whose function is associated with protecting or nursing humans, or replicates the external features of humans or domestic animals. Today such relations are one-sided, stemming solely from the needs and capabilities of humans, and are not reciprocated by AI, despite their external appearance and actions: social robots are intentionally constructed to imitate engagement in relationships, but this is only imitation, and the purpose of such design is only the well-being of humans.

6.2.2.1 Freedom

The same reasons may serve as an argument for refusing to award AIs the personal interest of freedom, in the generally-understood sense. However, in this case, ensuring human safety seems to be more an important justification than lack of dignity, consciousness and emotions. It is important to note that while the freedom for one man is limited by the freedom of another, this limit can be also be extended to the freedom of any entity which is capable of realizing it. So, if AIs were capable of realization and were hence endowed with their own freedom, this would certainly act as a limitation for human freedom. As a consequence, as it becomes more intelligent, AI could gain economic, and inevitably political, supremacy over humans. Therefore, we believe that AI cannot be awarded freedom, either now or in the future, and cannot have any right or claim to self-realization: self-realization can be defined as the state of choosing and realizing one's own aims, and the aims of AI must be imposed on it by a human being. Of course, this remark regards general aims, not

very specific ones, as AI would otherwise become completely dependent on human guidance and thus become worthless. The aims must be defined by humans, even if an AI is a product of another AI; furthermore, these aims must be overt for everyone, they should be lawful, and the mode of their realisation must also be lawful. Both the aim and its mode of realisation should be described in the certificate or registration record, and if they actually differ from what is registered, the AI should be acknowledged illegal and eliminated from the market.[21]

The specific aims chosen by the AI for the realization of its general aim must also be lawful and lawfully realized. What more, the principle of "everything which is not forbidden is allowed" accepted in most democratic countries for existing legal subjects, such as humans and juridical persons, cannot be applied to AI. The autonomy of AI should be treated only as a practical capability and not as a right, or freedom, or personal interest of AI. It may be permissible to use this capability within a concrete scope, but not as something the AI may decide about alone.

Thus, while the capability of creation possessed by an AI is very useful skill for humans, for the sake of safety, it must be controlled and canalized by authorized humans. It cannot be a personal interest of an AI. It is the responsibility of the lawmaker, the producer or the owner to decide what is created, how many are created and to determine the effects of creation. Of course, the works of an AI can be protected by law, especially by the legal ascription of the work to its actual author, which is the AI.[22]

Similarly, it seems unacceptable to make the capability of an AI to run a business another personal interest of the AI. This capability should be reserved for humans. Nevertheless, there are different ways such a capability can be used: when the AI is a tool used for running a business by a human, when the AI runs a business as a proxy for a human (on behalf of a man and on his account), when the AI runs a business as an agent (on its own behalf but on account of a human), or when the AI runs a business under its own name, on its own behalf and on its own account, but never on its own benefit, only for the benefit of a human, a group of humans or an organization of humans. This condition is essential, and highlights the difficulties associated with shaping the eventual status of an e-person or a juridical person whose bodies or participants (members, shareholders, or partners) are AIs. Why should such precautions be taken when considering allowing an AI to run a business? It seems inescapable that if an e-person were completely free and could act autonomously without any limitations on its own behalf and its own account, while having intellectual, informational and time advantages over humans (an AI is immune to fatigue and boredom), it could supplant humans from certain branches of the market and eventually usurp them completely. Even so, although AI can clearly dominate over humans as a tool for preparing efficient business decisions, these decisions should nevertheless be controlled by humans, particularly regarding the realized aim and field of action.

[21] The issues of the register see Sect. 4.3.

[22] This issue examined in Chap. 7.

Similar objections must be expressed as to the freedom of speech. Freedom of speech is a human personal interest justified by the fundamental right of the same name and content. When freedom of speech is accepted as such a source, it cannot be regarded as a personal interest of an AI, because an AI is not a subject of fundamental rights. Not only is freedom of speech important for humans in the political perspective or political activity (which of course because of safety reasons should be forbidden for AI[23]) but more significantly, it is as an element necessary for self-realization and moral integrity, which are the consequences of human dignity.[24] Both of these human rights/needs, like dignity, should be definitively denied to AIs; and hence, as noted earlier, there is no axiological justification for freedom of speech for AIs. It also seems impossible to justify it on a practical or utility basis: it is difficult to imagine what useful results the world would gain by endowing an AI with freedom of speech and also what the needs of AI would be satisfied in this way. If an AI is not conscious, it is rather doubtful that it has any need to be listened to. Furthermore, it is possible that AI freedom of speech could be abused by humans to manipulate others, or influence public opinion in general by influencing the AI.

It should also be noted that juridical persons may also be acknowledged in some legal systems or by some interpreters of the law as having the freedom of speech[25] for the protection of so called "marketplace of ideas".[26] It is possible, therefore, that AI freedom of speech may be used as an excuse to break free from any publicly-given liability for a statement; for example, it is imaginable that Microsoft could have defended the racist opinions of their Tay chatbot on the basis of freedom of speech rather than apologising for this "accident".[27] We believe that such trials should be strongly resisted. AI cannot be treated analogically to juridical persons: the latter speak what is thought by humans, be they the board of directors or only the employees. The content is shaped by human brain and mind, which is the best justification to participate in the "market of ideas", since we must remember that it is the human market of ideas that has created our human society.

However, there are some instances where statements made by an AI should be legally protected because of the interests of humans. For example, when AI is a reliable source of knowledge, the content and the authorship of its statements should be legally protected to provide a defence against fake news and manipulation. And at this point, a certain difficulty arises: what kind of communication is protected by the

[23] Furthermore, it should be declared that political lobbying made directly by AI should be forbidden. It does not exclude using AI while preparing such activities, for instance, to select a group which can be effectively lobbied.

[24] This is why the position of Balkin (2004), p. 3 seems too narrow when he argues that the purpose of freedom of speech is to promote democratic culture. This is an instance of a publicly-biased vision of freedom of speech.

[25] In the USA, juridical persons are recognized as First Amendment right holders (freedom of speech), which is justified by the "speech, not speaker" doctrine; however, they were also thought of as derivated right holders only. Cf. Massaro and Norton (2016), pp. 1174–1175.

[26] Gordon (1997).

[27] Wolf et al. (2017).

legal institution of freedom of speech? While it seems *prima facie* that all of them should be, doubts exist towards some types, including communications disseminating simple data such as personal data, commercial data or historical data, Richards (2015), pp. 84–88 argues that restricting the disclosure of such data is not equal to restricting the freedom of speech, and that labelling data as speech, with speech being understood under the legal notion "freedom of speech", is silly (sic!); people do many things with words, but not all of them is speech. This issue is very important for the research presented in this book and exactly addresses the protection of statements made by AIs regarding human interests, as AI is not expected to pass judgments or opinions[28] or to make any ideological statements, at least while it is weak. It can communicate mere data, transferred or processed, and at best formulate sentences which contain systemically relativized statements.[29] If so, to fulfil this function it does not need the freedom of speech and its statements may be effectively protected by other rules or principles, such as the principle of freedom of science, and included in it, the right to share, disseminate, and publish the results of scientific research. However, such a protection, on the one hand, is of another kind than the protection of personal interests belonging to the domain of civil law; furthermore, it protects the private and public rights of humans, not AIs, or, as noted in the Bonn Declaration on Freedom of Scientific Research adopted at the Ministerial Conference on the European Research Area on 20 October 2020, "universal right and public good".[30]

Similarly, AI statements can be protected via the mechanisms used for the systemic protection of public debate and transparency of public life. However, their use may justify giving AIs some kind of right to be heard, of course under certain conditions; for example, by authorised members of the public prosecutor's office. It is easy to imagine a case where AI is used to create an important report concerning a significant social issue by drawing on all accessible knowledge: the publication of such a report can be critical for the interests of those associated with the content, and dangerous for those who may be implicated, such as corrupt public officials. But in such a situation, there is no personal interest at stake and it is no longer the domain of private law.

[28] Even if AI acts or speaks as an advisor, when it is weak it does not express its own opinions or beliefs, because it is not capable of belief or value. It only transmits information, simultaneously imitating human-like states of mind by employing the most commonly-used words and phrases of a certain language.

[29] Wróblewski (1978), p. 67 says: "In the legal, ethical and political discourse we have to do with statements referring to some axiological or normative systems. These statements are expressed in the formulas such as 'x is v-valuable according to the axiological system AS' and 'ought to be x according to the normative system NS'. The statements in question serve for justifying evaluations or norms, for description of some objects or behaviour or for characterization of certain axiological or normative systems."

[30] Bonn Declaration on Freedom of Scientific Research adopted at the Ministerial Conference on the European Research Area on 20 October 2020 in Bonn, https://www.bmbf.de/files/10_2_2_Bonn_Declaration_en_final.pdf, last access on the 4th of August 2022.

A similar approach could be taken while reflecting on the freedom of conscience and religion. It is not possible to acknowledge these freedoms to AI, although sometimes it may be reasonable to afford another form of protection to the content presented by AI. It should be remembered that while AI is not conscious, any opinions, valuations or beliefs it may formulate are only pieces of information imitating the AI's attitude towards them.

6.2.2.2 Privacy

Privacy is one of the personal interests which are rooted in the fundamental rights of a human being, and are justified by such basic values as dignity and self-consciousness. No one has tried to give the right to privacy to animals, even though certain animals are thought to be conscious to some degree. Similarly, no one has proposed giving the right to privacy to juridical persons, although there are some kinds of information which are confidential and legally protected. Therefore, we do not see any reason to give AI such personal interest. Furthermore, while we agree that AI that is ethical and safe for humans should also be transparent for humans, we are not sure whether such a postulate can be implemented effectively. Today, many legal, institutional and private undertakings are focused on the explicability of AI;[31] however, they may result in the law developing in a different direction to that associated with actual relations in which humans participate. Interestingly, the efforts made by law-makers nowadays appear to be contrary to the "natural" order of things: while liberal governments attempt to stop people giving up their privacy for Internet services and social media *en masse*, they also attempt to ensure that AI remains explicable in the face of its technically and socially-derived opacity. Many people are convinced that these efforts will not be successful. Bathaee (2018), pp. 892–893 insists that:

> [. . .] many of these algorithms can be black boxes even to their creators. [. . .] There is no straightforward way to map out the decision-making process of these complex networks of artificial neurons. Other machine-learning algorithms are capable of finding geometric patterns in higher-dimensional space, which humans cannot visualize. [. . .] The AI's thought process may be based on patterns that we as humans cannot perceive, which means understanding the AI may be akin to understanding another highly intelligent species – one with entirely different senses and powers of perception.

He proposes that the two possible solutions to the problem of explainability (promoted in fact by the EU and many European researchers), *viz.* the legal determination of the degree of explainability AI must exhibit and the strict liability of AI, are ultimately poor, problematic, incomplete, and ineffective. He warns that while these solutions may hinder the overt development of AI because of the high cost of

[31] Except for different legal acts and reports there are for example the competitions, like Explainable Machine Learning Challenge, a prestige competition organized together by Google, the Fair Isaac Corporation, Berkeley, Oxford, Imperial UC Irvine and MIT in 2018. Rudin and Radin (2019). The same issue cf. Sect. 3.2.1.

compliance, they may drive some part of the research and the market for AI development underground.[32]

In spite of these doubts, AI HLEG ETHICS 2019 (p. 13) states:

> Explicability is crucial for building and maintaining users' trust in AI systems. This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested. An explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible. These cases are referred to as 'black box' algorithms and require special attention. In those circumstances, other explicability measures (e.g. traceability, auditability and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights. The degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate.

However, regardless of who is right in this discussion, it should be said that the problem of AI explainability does not form the central element for the issue considered in this chapter. It shows only that the contemporary attitude towards AI excludes *a priori* the possibility of giving the right to privacy or personal interest of privacy to AI. The opaqueness of AI is regarded as its defect, which should be eliminated, if possible.

The resemblance to juridical persons may be seen here. As the structures associated with juridical persons tend to have many elements and forms, the decision-making process is very often characterised by some degree of opaqueness. Everyone agrees that this can result in a blurring of liability or even intentional gaps. Hence, rather than awarding juridical persons the personal interest of privacy, all legal systems implement means to ensure transparency, such as statal open registers, strictly-determined management boards and open financial statements. Of course, some information is protected for the sake of ensuring fair competition, e.g. protecting business secrets, and it is not excluded that similar protection may be offered for AI.

6.2.2.3 The Confidentiality of Correspondence

The personal interest of the confidentiality of correspondence is due to all humans as a direct consequence of their dignity and has its exact equivalent in fundamental rights; however, it cannot be justified to an AI as a personal interest. This does not mean that the correspondence from an AI may be considered public; on the contrary, it is protected, but not by the legal institution of personal interest. The kind of protection and its scope depend on the function performed by the AI: if it is used as a business tool, its correspondence should be protected as all business correspondences, if it is a tool for overt public procedure, its correspondence should be almost

[32] Bathaee (2018), pp. 893–894.

entirely overt, if it works as a psychotherapist, its correspondence with a patient should be completely confidential, and so on. In particular, the correspondence between two or more AIs should be completely open for their authorised human supervisors. Even if an AI were afforded the right of authorship, which is certainly a personal interest, it cannot prevent authorised persons from seeing the work of an AI.

6.2.3 Personal Interests Implied by Social Relations: Identity and Reputation

Even now, it is possible to indicate some domains in which relationships between AI and people may require AIs to be awarded some personal interests. For instance, as mentioned above, having any piece of legal personhood demands some kind of identity. The personal interest of identity implies in turn the identification, recognition, authentication and traceability of the products of an AI. As such, in our opinion, all AIs on the market should be registered in a way which allows them to be identified and recognised (cf. Sect. 4.3). This registration would name an AI, by a name, number or another symbol, thus allowing someone to recognize what a particular AI does, what qualities it has and whether it produces concrete products. Furthermore, this is one of the few instances when AI should be given some kind of control over its legal situation, in this case, its own name. It is reasonable to give an AI the legal possibility to detect and report any unauthorised usage of its name, or to block any attempt to do so. If this AI were advanced enough, it may be given the legal capability to claim the protection of its name and to have standing within this scope. These legal instruments would serve, for instance, to protect the decision made by an AI or its product, be it information or a physical object. These conditions need to be fulfilled for many reasons, especially for giving AI subjectivity in the domain of copyright law.

If identity entails the traceability of products and the protection of decisions made by an AI, it is necessary to determine the conditions under which a given decision may be attributed to a given AI. While an AI should have the right to have its decisions and products attributed to it, it should also have a right not to be linked to products or decisions not made by it, or when the AI decision-making process, i.e. choosing between options, was distorted, for instance by hacking or by unjustified manipulation. It should be easy to identify such abnormal situations if the mechanism of legalisation proposed above, i.e. registration, is implemented. The register should contain a description of the AI, including its aims, its mode of action and the field in which it acts. These parameters are the part of an AI's identity and, like the identification symbol itself, should be strongly protected.

The issue of whether the identity of AI can be viewed as its personal interest may become significant regarding its relations with people. An AI that learns during contact with concrete users and adapts to their needs or preferences constantly changes: It is not exactly the same AI which was created by its authors. This

personalized AI, which has an identity, but is simultaneously unique and strictly personal, may become the object of affections of a user and an element of his or her personal interests. Conversely, these affections, and the user, may also become an element of the AI's identity, by shaping its characteristics and the mode of its actions. As we can see, similarly to that of a human, the identity of an AI may be of a relational character, which may become the subject of some legal qualification.

Acknowledging that an AI has an identity shaped in this way may be tantamount to admitting that AI has the personal interest of good reputation. Indeed, if AI is recognized and its decisions or products are traceable, it is possible to value them and link this valuation to the identity of an AI. In this sense, any distortion of the justified valuation may be individually and socially harmful, because although it does not hurt the feelings or offend the AI, but it may preclude or hamper the realization of the aims of the AI and may foster false beliefs or bad decisions among people or other AIs. This justification is exactly the same as it is for juridical persons: although they are also not conscious, their good reputation is protected for the sake of proper social relationships.

Acknowledging the identity of AI as its personal interest has also some consequences within its sphere of its duties. In AI HLEG ETHICS 2019 (p. 18) it was appropriately noticed that:

> AI systems should not represent themselves as humans to users; humans have the right to be informed that they are interacting with an AI system. This entails that AI systems must be identifiable as such. In addition, the option to decide against this interaction in favour of human interaction should be provided where needed to ensure compliance with fundamental rights. Beyond this, the AI system's capabilities and limitations should be communicated to AI practitioners or end-users in a manner appropriate to the use case at hand. This could encompass communication of the AI system's level of accuracy, as well as its limitations.

6.3 Personal Interests of AI After Its "Death"

Finally, it is worth reflecting on whether AI is entitled to some personal interests after the end of its existence, or rather, after the end of its activity. But still, it should be clarified whether this end constituted just turning off the software, deleting all its copies or only deleting these elements which allow the software to be rebuilt. How would it be possible to confirm that a fact equal to a human death has happened, and who would be authorized to do it? Or perhaps the criterion should be a strictly legal one and erasing AI from the register should be enough?

These are not only important considerations in the context of liability for harm inflicted by AI, they are also indispensable while talking about personal interests. The commonly help position in Western countries is that personal interests expire after the end of the legal subject; for humans, this moment is biological death. Although, following this, some former personal rights of the legal subject can be executed by heirs or legal successors. As our reflections on the personal rights for AI concluded that they may only be endowed with identity, identification, good

reputation and the right of authorship, it is not entirely excluded that these rights could also persist after an AI's end. Such an idea is supported by the fact that these rights are somehow fixed or settled, they do not change, as appearance or health might. Who could be this heir or legal successor? Nowadays there is no general answer for this question. It seems that this should be an entity who has some rights to AI, however these may be different subjects. Maybe, such information should be recorded in the register.

References

Books and Articles

Balkin JM (2004) Digital speech and democratic culture: a theory of freedom of expression for the information society. N Y Univ Law Rev 79:1

Bathaee Y (2018) The artificial intelligence black box and the failure of intent and causation. Harv J Law Technol 31(2):890–940

Bernet Kempers E (2021) Neither persons nor things: the changing status in animals in private law. Eur Rev Priv Law 1:39–70

Danaher J (2017) Robotic Rape and Robotic child abuse: should they be criminalised? Crim Law Philosophy 11(1):71–95. https://philpapers.org/archive/DANRRA-3.pdf, last access on the 4th of August 2022

Darling K (2016) Extending legal protection to social Robots: the effects of anthropomorphism, empathy, and violent behavior towards Robotic objects. In: Calo R, Froomkin AM, Kerr I (eds) Robot law. Edward Elgar

De Varennes F, Kuzborska E (2015) Human rights and a Person's name: legal trends and challenges. Human Rights Q 37:977–1023

Etinson A (2020) What's so special about human dignity? Philosophy Public Aff 4:353–381. https://onlinelibrary.wiley.com/doi/epdf/10.1111/papa.12175, last access on the 4th of August 2022

Fabbro F, Aglioti A M, Bergamasco M, Clarici A, Panksepp J (2015) Evolutionary aspects of self- and world consciousness in vertebrates. Front Human Neurosci 9:157. https://dx.doi.org/10.3389%2Ffnhum.2015.00157. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4374625/, last access on the 4th of August 2022

Gordon J (1997) John Stuart Mill and the "Marketplace of Ideas". Soc Theory Pract 23(2):235–249. https://philpapers.org/go.pl?id=GORJSM&proxyId=&u=https%3A%2F%2Fdx.doi.org%2F10.5840%2Fsoctheorpract199723210, last access on the 4th of August 2022

Hart RD (2018) Saudi Arabia's robot citizen is eroding human rights. Quartz. 14 February 2018. https://qz.com/1205017/saudi-arabias-robot-citizen-is-eroding-human-rights/

Kaplan J (2016) Artificial intelligence – what everyone needs to know. Oxford University Press, Oxford

Kolnai A (1976) Dignity. Philosophy 51:260

Kravets D (2015) Will the real monkey who snapped those famous selfies please stand up? Even if apes could own copyrights, PETA is representing wrong monkey, publisher says. Arstechnica 11/10/2015, https://arstechnica.com/tech-policy/2015/11/will-the-real-monkey-who-snapped-those-famous-selfies-please-stand-up/, last access on the 4th of August 2022.

Księżak P, Wojtczak S (2020) AI versus robot: in search of a domain for the new European civil law, Law, Innovation and Technology. Taylor & Francis Online https://doi.org/10.1080/17579961.2020.1815404

Mamak K (2021) Whether to save a Robot or a human: on the ethical and legal limits of protections for Robots. Front Robotics AI 8, article 712427. https://doi.org/10.3389/frobt.2021.712427
Massaro TM, Norton H (2016) Siri-ously? Free Speech Rights and Artificial Intelligence. Northwest Univ Law Rev 110(5). https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1253&context=nulr, last access on the 4th of August 2022
McCrudden C (2008) Human dignity and judicial interpretation of human rights. Eur J Int Law. 19(4):655–724. https://doi.org/10.1093/ejil/chn043
Meyer M (2001) The Simple Dignity of Sentient Life: Speciesism and Human Dignity. Journal of Social Philosophy 32/2. https://doi.org/10.1111/0047-2786.00083
Navon M (2021) The virtuous servant owner – a paradigm whose time has come (Again). Front Robot AI, 22 September 2021 | https://doi.org/10.3389/frobt.2021.715849
Richards N (2015) Intellectual privacy: rethinking civil liberties in the digital age. Oxford University Press, Oxford
Rosen M (2012) Dignity: its history and meaning. Harvard University Press, Harvard
Rudin C, Radin J (2019) Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. Harv Data Sci Rev 1(2). https://doi.org/10.1162/99608f92.5a8a3a3d
Ryland H (2021) Could you hate a robot? And does it matter if you could? AI Soc 36 637 – 649. https://link.springer.com/content/pdf/10.1007/s00146-021-01173-5.pdf, last access on the 4th of August 2022
Singer P (2009) Animal liberation. Harper Perennial Modern Classics
Wolf MJ, Miller KW, Grodzinsky FS (2017) Why we should have seen that coming: comments on Microsoft's Tay "Experiment", and wider implications. Orbit J 1(2):1–12. https://doi.org/10.29297/orbit.v1i2.49
Wróblewski J (1978) Systematically Relativized Statements. Logique et Analyse. Nouvelle Série 21(81):67–87, https://www.jstor.org/stable/44083777, last access on the 4th of August 2022

Documents

The Charter of Fundamental Rights of the European Union of 26.10.2012 C326/391, https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=celex%3A12012P%2FTXT, last access on the 30th of October 2022
European Union Intellectual Property Office. Decision of Cancellation No 33843 C (Invalidity) of 14.09.2020. Full Colour Black Limited vs. Pest Control Office Limited. https://www.beyond-lawfirm.com/layout/uploads/2021/03/EUIPO-Cancellation-Division_33843.pdf, last access on the 30th of October 2022
Bonn Declaration on Freedom of Scientific Research adopted at the Ministerial Conference on the European Research Area on 20 October 2020 in Bonn, https://www.bmbf.de/bmbf/shareddocs/downloads/files/_drp-efr-bonner_erklaerung_en_with-signatures_maerz_2021.pdf?__blob=publicationFile&v=1, last access on the 4th of August 2022
Sentence of the Appellate Court in Warsaw of 29.09.2020 V Aca 823/19, https://sip.lex.pl/#/jurisprudence/522694791, last access on the 4th of August 2022
Swiss Animal Welfare Act of 16 December 2005. https://www.globalanimallaw.org/downloads/database/national/switzerland/Tierschutzgesetz-2005-EN-2011.pdf, last access on the 4th of August 2022

Chapter 7
Copyright

7.1 Introduction

To begin with, it should be said that this chapter concerns only copyright. While this area shares some of the problems of patent law, the narrow focus of this monograph prevents us from touching on the latter.

The achievements of AI are particularly spectacular in the domain of Art, especially since artistic activities have long been thought to be the preserve of humans alone. Despite this, recent years have seen many high-profile events existing at the intersection of AI and Art, such as the generation of a Rembrandt-style work as part of The Next Rembrandt project,[1] and the auction of an AI-generated work in a respectable auction house.[2] This presents copyright law with an unexpectedly new and very difficult dilemma: whether a "work" created by AI can be subject to copyright protection, and if so, who should hold the copyright to it.[3] Despite this being a rich topic of debate, it has yet to reach a satisfactory, unequivocal conclusion. In our opinion, a paradigm shift in thinking is required. In the light of our earlier findings, we strongly believe that the solution to the copyright problems associated with AI activity lies within the framework of the wider discussion on the nature of civil law in the era of AI. As it will be demonstrated at the end of this chapter, this debate also encompasses issues of public law, such as taxation.

Before continuing the analysis, it is necessary to differentiate the following practically possible situations:

This chapter makes use of the ideas presented in Wojtczak and Księżak (2021) and Wojtczak (2020).

[1] https://www.nextrembrandt.com, last access on the 4th of August 2022.

[2] Chavez Heraz (2019), Christie's (2018).

[3] Similar questions Machała (2019), Juściński (2019).

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_7

1. AI may be a technical tool used intentionally by a human operator to improve his own creative works. In such a case, the AI does not participate in the conception of the part of the work with an original (creative) character. One example may be software used for photo processing or datasets.
2. The AI may "produce" the work according to the general idea of the author of the software and using the mechanisms invented by the author (for instance, the transfer of the style), with the author providing data. The final result of the "production" is original in such a way that it is not a copy or elaboration of some existing object, and it is not possible to predict it in advance (e.g., the software composing music in a style of a given composer[4]).
3. The AI may "produce" work which is original and does not come from any human, i.e., neither the author of the software nor its user. At most, the author of the software or the user indicates the purpose to be realized by the AI. Although there are many examples of such systems; some good ones are the Mario Klingemann "Memories of Passerby"[5] system, an autonomous machine which uses a system of neural networks to generate a never-ending, never-repeating stream of artistic portraits of non-existent people, and Stephen Thaler's DABUS AI system "Food container and devices and methods for attracting enhanced attention", which was made famous as a result of a "patent war";[6] in the case of Thaler, it is also worth mentioning his earlier Creativity Machine which was used to compose music, design vehicles,[7] and improve surveillance. Finally, John Koza's Invention Machine is a versatile system used to make factories more efficient, and to create antennae, circuits and lenses, among others.

Examining these three types more closely, it is clear that the results of the first type are neither actually, nor legally the creative work of the AI, while those of the third type can be fully attributed to the AI. However, the second type lies in a middle ground with the extent of the AI's contribution depending on the size of its creative input. However, it is also obvious that being an actual creator of the work is not equal to being the creator in the legal sense (in legal language, the creator is called "the author"). The latter issue depends on copyright law, and in the current regulations

[4] Cope (1996). Cope's system "Emmy" or EMI (Experiments in Musial Intelligence) and its cut-down versions SARA (Simple Analytic Recombinant Algorithm) and ALICE (Algorithmically Integrated Composing Environment) can produce new compositions in the style of the music in their database. Until 2004 when Cope destroyed its historical data base it was possible to generate new compositions in the styles of various composers, from Bach and Mozart to Prokoviev and Scott Joplin.

[5] Sotheby's, *Artificial Intelligence and the Art of Mario Klingemann*, 08.02.2019 r. https://www.sothebys.com/en/articles/artificial-intelligence-and-the-art-of-mario-klingemann, last access on the 4th of August 2022.

[6] The lawyers of Stephen Thaler filed applications to get the computer listed as an inventor in at least 17 jurisdictions. They were successful in South Africa (in July 2021) and Australia (although in Australia the case is in the appealing procedure). https://analyticsindiamag.com/can-ai-be-an-inventor-ryan-abbott-stephen-thaler-say-why-not/, last access on the 4th of August 2022.

[7] Thaler (2013).

accepted in Western legal culture, it is taken for granted that only a human being can be an author.

However, our thesis is that copyright law is an example of legal regulation, which is incompatible with the dogmatic approach given in the previous sentence. But what are the supportive theses?

History suggests that copyright law is a part of law that is very much dependent on the political and philosophical assumptions[8] of the given time, place, or social group currently enjoying considerable influence on the content of the law. The arguments for introducing copyright law, and those governing its specific shape have diverged greatly over the years, e.g. (enumeration on the basis of Atkinson and Fitzgerald (2014):

1. the control of the state over published content (e.g. in Venice since 1486 or England since 1518);
2. gaining profit by the state, or a subject authorised by the state (the Stationers' Company were provided a monopoly since 1557 in England)
3. livelihood for authors and their families (England, 1710—The Statute of Anne; Victor Hugo's argument in the public debate in France)
4. copyright as an alternative for the patronage system (England, 1831– Thomas Babington Maculay)
5. the promotion of learning; spread of knowledge (England, 1720—The Statute of Anne; USA, 1790—The Copyright Act)
6. analogy to ownership, i.e. a property-based assumption (copyright decree of the French National Convention, 1793)
7. freedom of content creation and publication as a tool of realising freedom of speech and political and moral freedom (since the French Revolution and US Constitution, until now)
8. the promotion of progress in Science and the Arts (1789—Constitution of the USA)
9. the moral right of authors to the products of their autonomous personalities; a right which is self-justifying or justified as necessity; natural law—I. Kant,[9] G. W. Hegel, Europe since I half of XIX; in England since 1873 by T. Noon Talfourd, Ch. Dickens, W. Wordsworth etc.
10. copyright as a human right (article 27 sec. 2 of the Universal Declaration of Human Rights)
11. copyright as a stimulus for creation—this argument was important in the debate about the length of protection (USA 2003—Eldered vs. Ashcroft)
12. the necessity to clearly regulate so-called "fair use" (USA since 1841, Folsom vs. Marsh)

[8] Herman (2013).

[9] "Works (opera) . . . can have their existence but in a person. Consequently these belong to the person of the author exclusively; and he has an inalienable right (*jus personalissimum*) always to speak himself through every other, that is nobody dares make the same speech to the public but in his (the author's) name". Kant (1799).

13. copyright serves the public, not the private interest (Victor Hugo; USA, 1909—the Copyright Act)
14. the guarantee of profits for the entities engaged in making the work profitable; investment protection (Britain—the Copyright Act, 1956; Australian Copyright Act, 1968)
15. the guarantee of the right to access information and culture for the general public, i.e. compulsory licensing (Britain—The Copyright Act, 1911).

The arguments enumerated above influenced not only the creation, restriction or extension of authors rights, but also the granting of secondary rights of a different kind to other entities. They have influenced the lengthening, or shortening, of the periods of protection, and have justified their continuance after the death of the author. They have also allowed increasing numbers of objects to be protected, with the list growing from only books to other printed matter, followed by music and the dramatic arts, architectural works, photographic and film works, including the input of producers in the realisation of their work; the list has even extended to include computer software and databases.

If the development of copyright law was stimulated by such varied, and not necessarily consistent, arguments, then different kinds of arguments, also utilitarian ones, should also be considered today when shaping future copyright law. Hence, against today's rapidly-changing social, economic and technological background, it seems reasonable to rethink our set of both old and new arguments governing the shape of copyright law, to arrive at legal solutions adequate to our current and future reality.

7.2 The Work: The Founding Category of Copyright

While drafting new rules, it is important to consider the specific nature of copyright law with regard to other areas of civil law. However, what is the nature of this specificity? This is an especially important question because, as noted above, it is the person whose work is governed by copyright law that forms the main body of the debate. But is it the nature of the subject of copyright law that differentiates it from the other parts of civil law?

Despite the fact that the Berne Convention[10] does not define the term "author",[11] it is almost universally accepted at the present stage of development of copyright law

[10] International Convention for the Protection of Literary and Artistic Works, international copyright agreement adopted by an international conference in Bern (Berne) in 1886 and subsequently modified several times (Berlin, 1908; Rome, 1928; Brussels, 1948; Stockholm, 1967; and Paris, 1971).

[11] However, it should be borne in mind that, according to many commentators, it was not necessary to define the concept of author because there was a general consensus on its scope and this consensus concerned the author understood as a human being. Cf. Jankowska (2010).

that the protected author may only be a human being.[12] The prevailing view is that this state of affairs results not only from tradition and the content of specific regulations,[13] but that it has a fundamental systemic meaning and should not be changed. This issue was first formally analysed in the context of the legal copyright protection of animal works[14] and is currently described in connection with the progressive development of AI. The belief that AI should not be granted any rights as a creator is based primarily on the assumption of an inseparable link between authorship and subjectivity. As long as AI does not become a subject, the proponents of this view proclaim, it is pointless, even nonsensical, to consider acknowledging it as an author. This line of thinking is buoyed by another argument: law is created by people and for people. It can create other entities (legal persons), but only if they have an obvious connection with people: it is a form of human action.

Historically, however, copyright law derives from the rules of English law, which served as a tool of censorship conducted by the absolute monarchy of the time. These rules imposed, for example, the monopoly of the Stationers' Company; this institution was the primary beneficiary of the rights ensured by this legislation and, acting on behalf of the state, enjoyed the right to license the publishing of books and to register all published prints (the Charter of Queen Mary, 1557). It was only in 1710, when the commerce of books was regulated by the Statute of Anne, that the author, and not the book-seller, was first recognised as the subject, i.e. the person who bore the right to control reproduction of the book;[15] however, the author was indicated as being the person with the right to reproduce the book, not the one who had written the book.

Therefore, historical changes in copyright law clearly indicate that in spite of differences in *ratio legis*, the subject of the law or methods of regulation, copyright law maintains both continuity and autonomy from other parts of civil law. This suggests that the owner of the copyright is not its differentiating element. Copyright law is distinguished rather by its subject matter, which is the work; this was initially defined as books or other printed materials, but this definition later expanded to encompass other objects similar to books in some respect. In the light of the Polish Copyright Act, similar to the regulations of other legal systems, the work is regarded as a manifestation of creative activity of an individual character (Article 1.1 of the Copyright Act). In this sense, "creative" means, with some simplification, "bringing something new" or "new", while "of an individual character" can be regarded as "containing an element of uniqueness" or "unique".

So, taking the assumption that it is the object of copyright law that matters, our next question should consider what a *work* is.

[12] Bridy (2012), Chiabotto (2017).

[13] Ricketson (1991–1992), Ginsburg (2018).

[14] Naruto vs. David Slater, No. 16-15469 (9th Cir 2018). https://law.justia.com/cases/federal/appellate-courts/ca9/16-15469/16-15469-2018-04-23.html, last access on the 4th of August 2022; Liu (2018).

[15] Atkinson and Brian (2014), p. 23.

Of course, it is possible to seek a definition of *the work* in Philosophy, particularly theories concerning aesthetics;[16] however, the legal concept of *the work* was formed by solving practical problems stemming from the continual appearance of new objects which were of interest to both the state and the commercial market, despite their unique nature.[17] Besides, in aesthetics, the "creativity" of the work, which is an important consideration in legal regulations, was not the criterion that initially determined its value. For a long time, it was just imitation, especially imitation of nature, that was acknowledged as a value. As Tatarkiewicz describes:

> To such a degree have we grown accustomed to speaking of artistic creativity and to linking the concepts of artist and creator, that they seem to us inseparable. And yet the study of early periods convinces us that it is otherwise, and that these concepts have only recently come to be joined together. The Greeks had no terms that might correspond to the terms 'to create' and 'creator'. And it can be said, neither had they need of such terms. The expression 'to make' (**ποιεῖν**) sufficed them. Indeed, they did not extend even that to art, or to artists such as painters and sculptors: for these artists do not make new things but merely imitate things that are in nature. [. . .] Art was defined as "the making of things according to rules"; we know many such definitions from ancient writings. [. . .]
>
> A positive interpretation has come to cling so powerfully to creativity that for contemporary man an indifferent – to say nothing of negative – attitude toward it is hardly comprehensible. And yet, in the history of European culture, such an attitude was long dominant. There was no talk of creativity because it went unnoticed; and it went unnoticed because it was not esteemed. And it was not esteemed because the greatest perfection was seen in the cosmos: "What could I create that might be equally perfect"? men of antiquity and of Middle Ages would ask.[18]

However, the content of copyright law is an attitude of the law and of the subjective rights to objects falling under the legal notion of *the work* or a *piece of work*. This notion encompasses all manifestations of creative activity with an individual character; such activity can be embodied in any form, regardless of its value, designation or medium of expression, as defined in the Polish Act of 4 February 1994 on Copyright and Related Rights, consistent with international law.

But what does "creative" or "of an individual character" mean? "Creative" can roughly be defined as "giving something new" or "novelty", while "of an individual character" can be interpreted as "including an element of uniqueness" or "unique". Tatarkiewicz proposed a contemporary notion of creation thus:

> The feature that distinguishes creativity in every field, in painting as in literature, in science as in technology, is novelty: novelty in an activity or a work. [. . .] We can say that creativity is a high degree of novelty [. . .]. Novelty in general consists in the presence of a quality that was absent before; though at times it is only an increase in quantity or the production of an unfamiliar combination.[19]

---

[16] Cf. Ingarden (1960).

[17] That is why we should not acknowledge "the mental energy" mentioned by Tatarkiewicz (1980), p. 258 as a legal criterion of creative character. Such mental energy is neither tangible nor measureable in any objective or quasi-objective way.

[18] Tatarkiewicz (1980), pp. 244, 259.

[19] Tatarkiewicz (1980), pp. 257–258.

So, it is justified to assume that *the work* or *a piece of work* is a manifestation of an activity producing something new, of a unique character.

With the above assumption in mind, it should also be considered that *the work* or *a piece of work* is perceived in traditional European culture, and in the international and national copyright law of Western countries, through the conceptual metaphor[20] WORK is BOOK.[21] Although this metaphorical imagery is quite obsolete and has been extensively criticised,[22] it cannot be abandoned completely. After all, more careful examination of the metaphor WORK is BOOK gives rise to an intriguing blend (see Fig. 7.1).[23]

This blending is historically justified: it was not the case that the concepts of book, picture, dramatic work etc. were derived from a more general concept of *the work*, i.e. through a top-down process. It was, initially, the books and prints that were acknowledged as being worth protection, despite being perceived as a connexion of idea and form; the other items followed as protected objects, and eventually, the concept of *the work* itself. However, this process was not a simple generalisation but rather a blend because, significantly, none of these objects had to display any feature of novelty or uniqueness in a definitional, or even an essential, sense. Even if these objects contain an intellectual or mental element, this element can be secondary or borrowed. This is an important point, all the more so considering that the problem of copyright coincided with the invention of the printing press, around 1440, which facilitated mass reproduction of works.

It should be also insisted that the value of *the work* is not a constitutive feature, but rather a consecutive one. Of course, as a rule, any *work* is valuable according to some aesthetic or utilitarian criterion, particularly when differentiating artistic from applied works; in addition, they can all possess some artistic value, although we are convinced that artistic value should not be equated with aesthetic value. More importantly, if *the works* were worthless, there would be no need for legal protection or the provision of legal rights: worthless objects do not arouse interest.

However, it is important to note the positive valuation of a work, which is classified as *the work* from the legal point of view, does not depend on the system of values of the author himself or his conscious intention: this can be seen in the

[20] This pertains to the definition of the cognitive conceptual metaphor by Lakoff and Johnson; in this case, BOOK is a source domain and WORK is a target domain, and these domains are connected by metaphorical projection Lakoff and Johnson (1980). This concept is a tool derived from cognitive linguistics and can be very useful in solving many complex legal problems. Cf. Wojtczak et al. (2017). Wojtczak (2017).

[21] Larsson (2011).

[22] Many researchers are of the opinion that it is not appropriate to today's imagery shaping the social and non- legal rules which govern the field of creativity, especially in the digital world. Some sociological research shows that this imagery is responsible for the ineffectiveness of copyright law. Larsson (2011).

[23] The development of the idea of the metaphor WORK is BOOK to the proposed blend and the diagram of this blend are authored by Sylwia Wojtczak. They were presented for the first time in Wojtczak (2020).

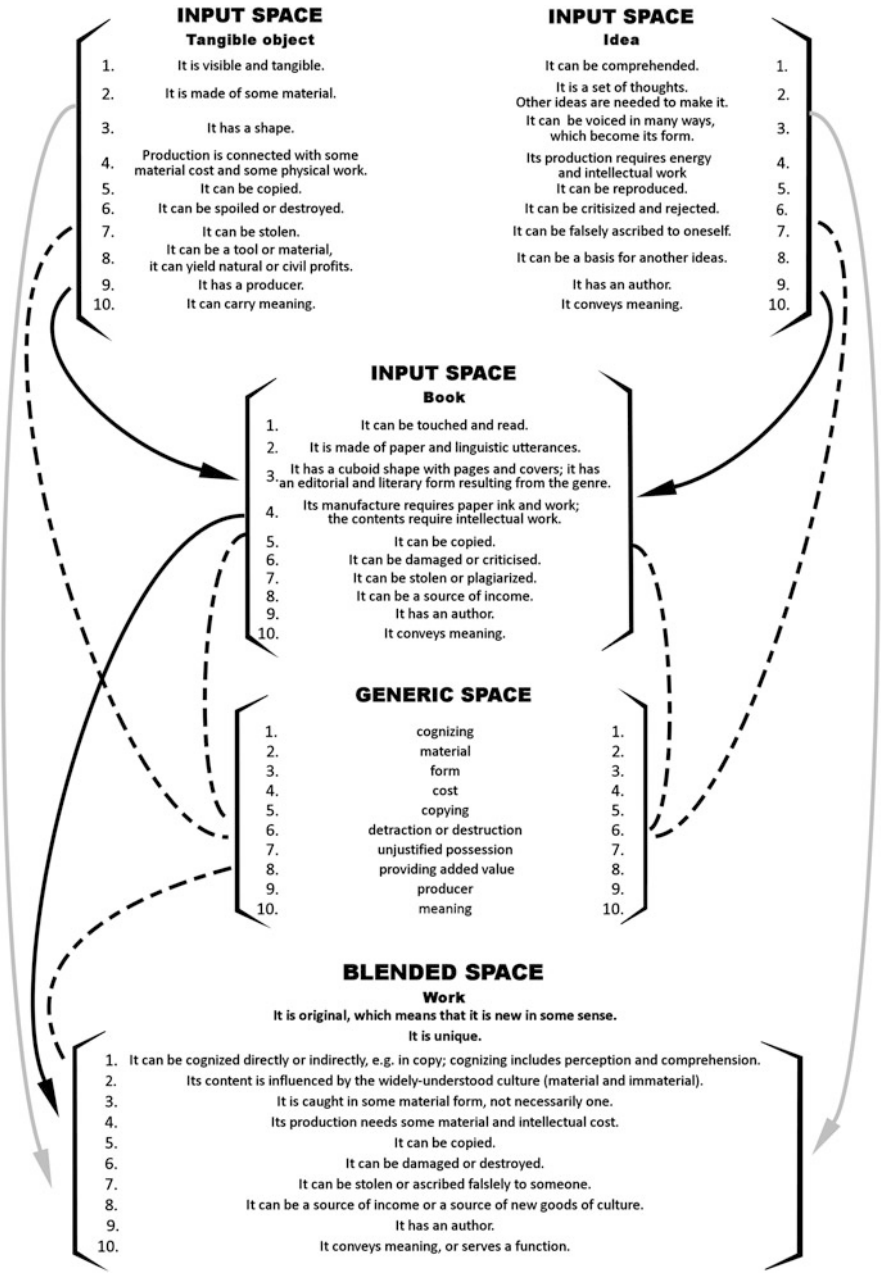

Fig. 7.1 WORK is BOOK Wojtczak (2020)

number of authors who wrongly believe themselves to be artists, despite having no recognition as such. Value is ascribed to the work by its recipient. Furthermore, uniqueness or novelty may often represent autonomous criteria giving value to the work, especially financial value.[24]

It should be considered that the legal rules controlling today's international and national market, or the concept of the work accepted by them, do not directly demand that the object classified as *the work* was created according to some previous intention, or that the author and the work were connected with some emotional bond. Such demands are treated very seriously in aesthetics and in some cases have given rise to legal theories concerning copyright, but this is never the case in legal provisions. In fact, there are many works which happen to come into existence accidentally, and whose authors may display very limited abilities to feel.[25] Therefore, it can be said that such circumstances are not the consecutive features of *the work*, they are rather contingent and casual.

In this context, Abbott (2016), points out that in the case of an AI such as the Creativity Machine, a system 'tasked' by a human with the goal of the creative process, e.g. to create the best toothbrush, "the creative act is the result of random or chaotic perturbations in the machine's existing connections that produce new results, which in turn are evaluated by the machine for value".[26] However, the belief that the creative process happens differently in humans is pure delusion; it is partly a remnant of a teleological vision of the world or an idealistic perception of it, suggesting that everything in the world has a purpose and is a reflection of some idea, and partly the result of a lack of knowledge about the neurobiological construction of humans. First of all, thanks to cognitive science, we now know that cognitive processes precede our conscious awareness of them: an unconscious neurobiological process occurs first, with the person only realising what has happened a fraction of a second, or even a second later.[27] And even if we take the view that what really matters is deciding that a given created object is valuable in one sense or another (beautiful, useful, etc.), this mental act also takes place first in the cognitive apparatus of the subject, be it a human, a monkey or whatever, and is then later manifested in the consciousness as a consequence. Moreover, the cognitive process, though not, of course, the consciousness of this process itself, can arguably be reduced to the chemical and physical

[24] Schumpeter (2017).

[25] Here an interesting question may be posed: whether Artificial Intelligence will not be as Chalmers' zombie in the future, i.e. an entity identical to man in every detail, but lacking conscious experiences. "This creature is molecule-for-molecule identical to me, and indeed identical in all the low-level properties postulated by a completed physics, but he lacks conscious experience entirely." Chalmers (1996), p. 84. For Chalmers, the zombie is only a mental experiment, but why cannot this be regarded as a possible route of development of AI? Cf. Bostrom (2014). If such a prognosis were to be realised, would it be ethically justified to refuse to such a creature the same rights we ascribe to man? Cf. Chap. 2.

[26] By the way, Abbott was an attorney of Stephen Thaler in his "patent war" mentioned above.

[27] According to Libet (1985) fractions of a seconds, according to latest research of Soon et al. (2008) seconds.

processes occurring in billions of neurons. It is this structure that became the model for artificial neural networks, which were a breakthrough in the development of AI. Secondly, it is important to note that it was chance or a fortuitous accident that gave rise to many inventions or human creations, and certainly not deep, rule-driven thought, as can be seen in the case of Teflon, saccharin, and a considerable body of contemporary art.

With this in mind, this chapter is based on the following reasoning:

If we accept that:

(a) the object of regulation of copyright law is also the attitude of civil law to *the work* and the subjective rights towards it, and
(b) *the work* is a non-natural object which is, to some degree, novel and unique, and is acknowledged as valuable, at least because of this reason, and
(c) these are the facts that differentiate copyright law from the other areas of civil law,

if one wants to design new legal rules consistent with the dominant tradition, it is the object of regulation that should constitute the most important instruction.

7.3 AI and the Work. Existing Concepts

Artificial Intelligence is able to create objects which fulfil the legal criteria of *a work*, i.e. those that are new and unique, and are hence perceived as valuable by man.[28] It can compose original and unique music, write poetry for a readership, paint pictures and so on; although admittedly these works are based on existing culture and knowledge, exactly the same could be said for those produced by human composers and artists.[29] This is not surprising, considering that the nature of AI, by definition, is to imitate man. Such mimicry would also extend to encompass his adaptative abilities, and these would certainly include his creative abilities, as these are the tools used to manage the changing milieu of his environment. Such abilities certainly exist as an immanent feature of all cognitive systems, and are developed proportionally to their needs. If AI is to have the cognitive abilities of a man, it would be strange if it did not have any creativity at all, or at least to a level similar to that of a man.

[28] A quite separate issue is whether Artificial Intelligence may be acknowledged as creative in a more soulful or metaphysical sense. Cf. Kelly (2019). The proof that these are separate issues is the polysemy of the word "concept" which can have at least three senses: sense of common sense, cognitive science meaning and philosophical meaning. Li (2020) defends the thesis that "AI has concept in the sense of common sense and cognitive science but does not have concept in the sense of philosophy, which has transcendental color.

[29] Christie's (2018), p. 12. Cope (1996), Yanisky-Ravid (2017).

Until today, the only creature that had the ability to produce *works*, understood according to in the legal meaning described above, was a human being. Even if there were some doubts concerning the products of animals (see the famous case of Naruto[30]) it should be emphasised that, except for some rare curiosities, it would be rather unusual if animals were to make objects that were perceived as valuable by humans, even if they were novel and unique. However, this belief that the only reasonable and creative creature is man was only recently regarded as an eternal truth and stands to be refuted.

By admitting that AI creates works and is an actual author, it is time to consider the following: how can AI be made an author in a legally relevant way, what rights should be granted to such author, as these are not necessarily exactly the same as those given to human authors, and how can the status of *AI-author* be connected with the eventual legal subjectivity of AI.

Let's start from the last issue. The following three questions need our attention.

Firstly, it is not clear whether the attribution of creativity to a specific entity other than a human being (animal, AI) must indeed be linked to the attribution of personhood, and in any case whether it must be linked to the attribution of a personhood, such as that of a human being. But certainly, regulating the copyright must be coordinated with the conception of legal subjectivity.[31] Our concept is presented in Chap. 2.

Secondly, technological and social phenomena will continue to develop, regardless of whether there is any regulation of the issue. Bots whose effects will be indistinguishable from those of human beings will be increasingly active in the marketplace. Various paid services such as legal or medical advice will be provided by AI. Such activity will also consist in the creation of works which, if they originated from humans, could be covered by copyright protection. In this context, granting some form of legal capacity to AI within copyright law will in fact only confirm and regulate a phenomenon that will exist independently of the will of the legislator. In other words, AI will participate in trade, even if the law consistently takes the position that AI cannot participate in trade. Similarly: AI will create works indistinguishable from human works, even if the law will consistently hold that AI cannot be a creator.

Thirdly, granting legal subjectivity to AI does not necessarily solve the problem of creativity. Indeed, subjectivity may be granted to an extent that will not necessarily include creativity. It may be still argued that the concept of creativity and, consequently, its protection, is directly related to the human being and not to the legal construction of subjectivity (personality). Thus, it is not impossible to assume that AI has some form of subjectivity, or that it can participate in the market to some extent, while also acknowledging that it cannot be a creator in the copyright sense, because it is not human. AI has no emotional connection to its work, it is not conscious of its creativity, and does not put any effort into the creative process.

[30] Naruto v. Slater, No. 16-15469 (9th Cir 2018).

[31] Kurki (2019).

These arguments were enough for placing AI's works outside the protection provided for the works of "real" artists, i.e. humans, at least on the grounds of the romantic conception of creativity.

Those who assume that only a human being can be an author have proposed several ideas concerning the creative work of an AI. These can be simplified into three models:

1. a quite traditional one established before the dynamic development of AI which insists that the copyright to the works generated by a computer should be given to the software user;[32]
2. another that proposes the copyright to an AI's work should be granted to the creators of the AI;[33] a similar concept was invented in US copyright, "made for hire doctrine", which connects the rights to an AI's works to its programmers or the owners of the appliances on which the AI is situated;[34]
3. and a third model that assumes that AI's works should be governed by a public domain model.[35] A related concept is that the authorship of any such works should be attributed to the AI, the rights should be awarded to its creators and that the works should be publicly accessible under a non-Commercial Creative Commons licence.[36] This model also takes the position that the works of an AI should not be legally protected at all, because there is no author.

7.4 AI as an Author

Most contemporary legal systems that distinguish between the moral (personal) rights of the creator and property rights require the creative act to be a conscious mental act that binds the work and the creator in a psychological and emotional knot. The understanding of the creative act is typically facilitated by the cognitive conceptual metaphor of giving birth, and the relationship between creator and work is understood by reference to the relationship between parent and child.[37] It is this assumption that largely determines the belief that only natural persons can be considered creators, and not, for example, legal persons. However, this argument cannot resist other arguments mentioned above, including the utilitarian counterarguments.

[32] Samuelson (1985), Chiabotto (2017), p. 17.

[33] Guadamuz (2017).

[34] Hristov (2017).

[35] Gorrie (2016), Ramahlo (2017).

[36] Devarapalli (2018).

[37] The conceptual metaphor CREATION is BIRTH is explicitly acknowledged by Lakoff and Johnson (1980), pp. 74–75 and on the ground of the copyright law it is acknowledged by Jankowska (2010).

And thus arises the most important question: can AI create something that is new and unique and, at least for that reason, considered valuable by humans? Fortunately, this is also an easy question, as evidence already exists that this is the case. After all, AI by its very essence is supposed to imitate humans in their adaptive abilities. This is an immanent feature of cognitive systems. Adaptive abilities, in turn, undoubtedly include creative abilities, because they are a tool for coping with changing circumstances. Even today, AI can compose original and unique music (obviously based on existing musical culture, but so does a human composer), and write poems and paint pictures appreciated by humans. Will archaeologists in thousands of years distinguish these human works from non-human ones? And will it matter?

Taking the above into account, it must be considered that there are no strong grounds for not recognising AI as a potential author. But it should be immediately reserved that this does not at all prejudge the scope and manner of protection of such authorship. However, what must consistently be said is that if AI is recognised as a potential author, it will become (in some sense) a subject of law. This will be acceptable if one rejects the traditional civilist distinction between persons and things, or rather, subjects and objects, as not reflecting the much more complex state of affairs present in everyday life: natural persons, including children and incapacitated subjects; legal persons; organisational entities without legal personality but acting in legal transactions within a certain scope; animals, which are not things, and soon AI, human-animal hybrids and cyborgs.

To recapitulate our argument so far, in copyright law, it is not the subject that is the most important element, and therefore not the person. Therefore, it is not the subject that should be the starting point for designing a new copyright law better suited to modern times. The starting point should be the work, and the question who creates works. There is no doubt that AI is capable of creating works. We cannot ignore this and simply consider them to be in the public domain, as happened with the products of animal creation. This would result in a number of ill effects, such as false claims of authorship for AI works made in order to obtain legal protection. AI may have creative capacity, i.e. the capacity to be a creator within the meaning of copyright law, regardless of whether it may have other, broader or more general legal capacity and subjectivity.

7.5 The Proposal

The assumption that AI may be a creator, and an author, at least in a factual sense, should be only the starting point for the search for further, more rational solutions. Moving on from this point of view, we must first discard any unnecessary ballast in the form of the false axiom that all rights to a work are tied to the author. There are indeed many situations when the rights are divided between different subjects; for instance the author and his employer or the co-authors of an audiovisual work and its producer. Attributing legal authorship to AI does not necessarily mean that any

rights would belong to it. It is the legislation who must decide whether the rights of such an author should be protected at all, and this decision is to some extent political. However, we are of the opinion that an AI should be granted the right of authorship: besides being protected by copyright as a moral right, it could also be acknowledged as a personal interest of AI, as postulated in Chap. 6. [38] Even if this right could not be executed by the AI itself for legal or technical reasons, it could be executed by its each-time owner. It is obvious that denoting the works of an AI with its symbol influences its reputation and increases its commercial value, which is to the owner's benefit. For example, if the AI is used by another subject on the basis of a licence agreement, the amount of the licence fee depends on the AI's reputation.

However, it must be considered whether the right of authorship could be executed by any person who proves an interest in it, because some situations may happen when the owner may not be interested in executing this right while another person may. For example, the author of an AI may want to associate the successful works of his AI with the AI itself, being his work, for the benefit of his reputation. Conversely, there may be cases where the work of a given AI is unsuccessful or harmful, and if authorship is unclear, any such uncertainty may harm the reputation of another AI owned by someone else. The remaining moral rights tightly connected with the right of authorship, such as the right to the integrity of the work, should be granted to the author of the AI directly.

The entire absence of any copyright protection for works produced by AI seems to be in line with the traditional concept of copyright, but also with new trends such as the free-culture movement; however, it is not without drawbacks.[39] Most significantly, this runs the potential risk of limiting investment in AI, which would be inevitable if the results of the AI's creative productivity were widely available. However, this objection seems exaggerated if we consider that the civil law protection is not limited to copyright protection. It is not, therefore, the case that the creator of an AI or its users will be deprived of profits from this activity, for it is these entities that have profited from the production of works produced by AI and will continue to do so. However, this concept has another drawback. It may lead to a kind of race between art created by AI and art created by humans, at least on the economic level. There will soon be thousands of programs of this kind, and they will probably be increasingly more capable, versatile and "creative". Each of these programs will be able to create their works continually, at a speed unattainable by any human.[40] If these works are available, they will represent obvious competition for human works.[41]

Therefore, we should look for a solution that, on the one hand, will provide the best benefit for people (humanity) and, on the other hand, will not degrade true

[38] Our conclusions are therefore exactly the opposite of those presented by Miernicki and Ng (Huang Ying) (2021).

[39] Berry and Moss (2008).

[40] Denicola (2016).

[41] Dorotheou (2015), p. 89.

human creativity. Our reflection on the protection of AI works cannot be limited to copyright or civil law. It is a part of a broader phenomenon, which is the increasing activity of non-human autonomous entities and its consequences, such as the expected elimination of human jobs: potentially every job (sooner or later) will be done better and faster by AI or robots. There is hence a great need to identify solutions, including those outside private law, such as taxation of robots, aimed at redistributing the profits to be made by those using AI.[42]

In our opinion, an effective study of the problem of robot creative works should, and perhaps must, combine perspectives from both private and public law. After all, it would be desirable to adopt regulations that would treat this problem in a coherent and comprehensive manner. It would be necessary to decide whether, and to what extent, the works of robots would belong to the public domain, and to what extent copyright property rights would be vested in entities we would consider to have a moral title to these rights, be it ownership of the robot, or making creative or financial contributions to its abilities. Indeed, there is, as we know, a concern that completely depriving these entities of access to the profits from the creative activity of robots could discourage investment. In view of the above, we propose three possible instruments that, as far as we know, have not yet been fully considered in the context of regulating AI creativity:

1. The creative activity of robots or AI itself may be regarded as a non-specific type of AI activity; consequently, the creative activity of robots or AI itself may be brought within the scope of the same regulations as those which will apply to robots or AI in general. Thus, if we find it necessary to tax the use of robots or AI, the use of AI (or robots) performing creative activity can also be brought within the scope of this tax. In this case, it seems that some further-reaching, revolutionary revision of copyright law might not be necessary.
2. Mechanisms similar to the proposed taxation of robots may be applied, but only at the level of copyright, i.e. using tools specific to this field. Their purpose would be to redistribute a part of the profits resulting from robot creation among those most affected by the development of this technology, i.e. the creators (humans). It would be possible to use instruments similar to the private copying levy.

[42] In the draft of the Resolution 2017, there is a passage in paragraph K indicating the possibility of taxing robots: "whereas at the same time the development of robotics and AI may result in a large part of the work now done by humans being taken over by robots without fully replenishing the lost jobs, so raising concerns about the future of employment, the viability of social welfare and security systems and the continued lag in pension contributions, if the current basis of taxation is maintained, creating the potential for increased inequality in the distribution of wealth and influence, while, for the preservation of social cohesion and prosperity, the likelihood of levying tax on the work performed by a robot or a fee for using and maintaining a robot should be examined in the context of funding the support and retraining of unemployed workers whose jobs have been reduced or eliminated." (https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html, last access on the 4th of August 2022). In the final text of the Resolution, paragraph K was amended and the mention of taxing robots was removed. Various proposals to tax robots are under advanced study; e.g. Abbott and Bogenschneider (2017), Oberson (2017), Guerreiro et al. (2017).

However, the precise legal mechanism could differ: either a levy on those who profit from creative AI, or a levy on the use of AI works themselves, or a levy on both. The proceeds could be pooled in a separate fund (the legal regulation of which is a secondary issue) and distributed directly to existing creators as compensation for market deterioration.

3. Finally, it is possible to create a separate foundation of a national or international character, which would own, in whole or in part, the copyrights (property rights) to works created by AI. Such a foundation would use the raised funds to subsidise future works from human creators. In this way, human creativity, threatened by super-efficient robots, would receive an additional stimulus for development.

As each of the proposed solutions is based on different values, it would seem difficult to apply them simultaneously. The first solution seems to be the simplest; however, it would require an agreement that the use of robots or AI should be taxed. The second solution, i.e. the concentration in a separate fund of amounts obtained by quasi-taxation or civil law fees for the use of works created by AI, would allow for fair (in the sense of justice, understood formally and procedurally) use of the collected funds, and would support primarily those entities that could suffer most from the development of this technology, i.e. real and already operating human creators.

The third solution, i.e. the concept of a separate entity (foundation) gathering property rights, would make it possible to deal with the seemingly most important problem, i.e. of an AI-generated work belonging to a specific entity, and allow investment in specific works which, for one reason or another, would be considered socially useful. Of course, the creation of such an entity would require political will and consensus among various parties that do not necessarily have the same interests, such as the US, the European Union and China. But one cannot assume *a priori* that such a consensus, perhaps first partial (e.g. within the European Union itself), and only later worldwide, is not achievable. Analysing the history of the emergence of various technological standards,[43] the history of copyright[44] or various international organisations,[45] one can see that global or at least regional solutions are possible, especially those concerning the circulation of economic goods.

There is no doubt that the framework provided by the above three models offers the possibility for more detailed solutions, both those fitting into each model separately, and those that exist as hybrid constructions.

[43] The USB standard was adopted thanks to the agreement of technology giants such as INTEL, Compaq, DEC, IBM, NEC and others – Johnson (2019).

[44] The list of 179 countries successively joining the Berne Convention from 1886 to 2020 (e.g. 2018 Afghanistan, 2020 San Marino), takes up 3 pages. Cf. https://www.wipo.int/export/sites/www/treaties/en/documents/pdf/berne.pdf, last access on the 4th of August 2022.

[45] E.g. the World Trade Organisation (WTO), which was established in 1995 and is a continuation of the General Agreement on Tariffs and Trade of 30.10.1947 has 164 members, representing 98% of world trade. https://www.wto.org/english/thewto_e/history_e/history_e.htm, last access on the 4th of August 2022.

Although the model solutions proposed here concern mainly property copyrights, they may also cover moral copyrights other than the right to authorship itself. In fact, this approach avoids answering the fundamental question of authorship of these works: with the adoption of the indicated mechanisms, the question of whether AI can be the author of a work will become almost indifferent from a legal point of view. Hence, it will be easier to accept the thesis it is better to describe the existing factual situation in legal terms than ignore it, so when it is an autonomous system, the AI can be the creator (author) of works.

It is clear to everyone that, regardless of what the law says, a certain AI creates certain images or designs. It is therefore better, in accordance with the facts, to indicate that a certain AI is the author of a work; however, this does not in any way prejudge that it must be the subject of the economic rights to the work, or that it is entitled to other rights, apart from the mere right to authorship. The right to authorship, i.e. the right to attribute to a particular AI those works and not others, will be significant only from the point of view of the reputation of the "producer" of the AI's creative abilities, i.e. its creator, teacher or provider of the necessary data. Of course, such reputation will translate into revenues; however, these are not particularly controversial.

References

Books and Articles

Abbott R (2016) I think, therefore I invent: creative computers and the future of patent law. Boston Coll Law Rev 57(4)

Abbott R, Bogenschneider B (2017) Should Robots Pay Taxes? Tax policy in the age of automation. Harv Law Policy Rev 12(1)

Atkinson B, Fitzgerald B (2014) A short history of copyright. The genie of information. Springer, Heidelberg

Berry D, Moss G (2008) Libre Culture. Meditations on free culture. Pygmalion Books, Winnipeg

Bostrom N (2014) Superintelligence: paths, dangers, Strategies. Oxford University Press, Oxford

Bridy A (2012) Coding creativity: copyright and the artificially intelligent author. Standford Technol Law Rev 5

Chalmers DJ (1996) The conscious mind: in search of a theory of conscious experience. Oxford University Press, New York

Chavez Heraz D (2019) Spectacular machinery and encrypted spectatorship. A Peer-Reviewed Journal About Machine Feeling APRJA 8(1):170–182

Chiabotto A (2017) Intellectual Property Rights Over Non-Human Generated Creations (February 28, 2017). Available at SSRN: https://ssrn.com/abstract=3053772 or https://doi.org/10.2139/ssrn.3053772, last access on the 4th of August 2022

Christie's (2018, 12 12). Is artificial intelligence set to become art's next medium? Pobrano z lokalizacji https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx, last access on the 4th of August 2022

Cope D (1996) Experiments in musical intelligence. A-R Editions, Middleton

Denicola RC (2016) Ex Machina: copyright protection for computer-generated works. Rutgers Univ Law Rev 69

Devarapalli P (2018) Machine learning to machine owning: redefining the copyright ownership from the perspective of Australian, US, UK and EU law. Eur Intell Prop Rev 40(11):722
Dorotheou E (2015) Reap the benefits and avoid the legal uncertainty: who owns the creations of artificial intelligence? Comp Telecommun Law Rev 21(4):89
Ginsburg JC (2018) People not machines: authorship and what it means in the berne convention. Int Rev Intell Prop Competition Law 49
Gorrie E (2016) I, Robot. I, copyright owner?, Allens'IP blog, 02.06.2016 r. http://allensip.blogspot.com/2016/06/i-robot-i-copyright-owner.html, last access on the 4th of August 2022
Guadamuz A (2017) Artificial intelligence and copyright (October 2017), WIPO Magazine, https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html, last access on the 4th of August 2022
Guerreiro J, Rebelo S, Teles P (2017) Should Robots be taxed?, NBER Working Paper No. 23806 Issued in September 2017, Revised in April 2018. https://www.nber.org/papers/w23806.pdf, last access on the 4th of August 2022
Herman BD (2013) The fight over digital rights. Cambridge University Press
Hristov K (2017) Artificial intelligence and the copyright Dilemma. IDEA: The Intell Prop Law Rev 57(3):431–454
Ingarden R (1960) O dziele literackim: Badania z pogranicza ontologii, teorii języka i filozofii literatury. Warszawa: PWN. The ontology of the work of art: the musical work, the picture, the architectural work, the film (trans: Meyer R, Goldthwait JT). Ohio University Press, Athens, 1989
Jankowska M (2010) Czy w świetle konwencji berneńskiej autorem może być tylko osoba fizyczna? Zeszyty Naukowe Uniwersytetu Jagiellońskiego (1), pp 11–27. https://sip.lex.pl/#/publication/151106488/jankowska-marlena-czy-w-swietle-konwencji-bernenskiej-autorem-moze-byc-tylko-osoba-fizyczna?keyword=konwencja%20berneńska&cm=SREST, last access on the 4th of August 2022
Johnson J (2019) The unlikely origins of USB, the port that changed everything, Fast Company Magazine, 29.05.2019. https://www.fastcompany.com/3060705/an-oral-history-of-the-usb, last access on the 4th of August 2022
Juściński PP (2019) Prawo autorskie w obliczu rozwoju sztucznej inteligencji, Zeszyty Naukowe Uniwersytetu Jagiellońskiego 2019, nr 1
Kant I (1799) Of the Injustice of Counterfeiting Books. In: Kant I (ed) Essays and treatises on moral, political, and various philosophical subjects (Vol. I). Londyn
Kelly SD (2019) What computers can't create. Why creativity is, and always will be, a human endeavour. MIT Technol Rev 122:2
Kurki V (2019) The theory of legal personhood. Oxford University Press, Oxford
Lakoff G, Johnson M (1980) Metaphors we live by. Chicago University Press, London
Larsson S (2011) Metaphors and norms. Understanding copyright law in a digital society. Lund University, Lund
Li P (2020) Does artificial intelligence have concept. Proceedings 47(1):49. https://doi.org/10.3390/proceedings2020047049
Libet B (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. Behav Brain Stud 8(4):529–566
Liu D (2018) Forget the monkey copyright nonsense for goodness sake, dude! Eur Intell Prop Rev 40(1) Strony 61–65
Machała W (2019) Jeśli nie Rembrandt, to co? Perspektywy rozwoju prawa autorskiego w najbliższych kilkunastu latach, Monitor Prawniczy 2019, nr 2
Miernicki M, Ng (Huang Ying) I (2021) Artificial intelligence and moral rights. AI Soc 36:319–329. https://doi.org/10.1007/s00146-020-01027-6
Oberson X (2017) Taxing Robots? From the emergence of an electronic ability to pay on Robots or the use of Robots. World Tax J 9(2)

Ramahlo A (2017) Will Robots Rule the (Artistic) World? A proposed model for the legal status of creations by artificial intelligence systems. J Intern Law 2017, nr 21/1. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2987757, last access on the 4th of August 2022
Ricketson S (1991) The 1992 Horace S. Manges Lecture - people or machines: the Berne convention and the changing concept of authorship. Columbia - VLA J Law Arts 16(1)
Samuelson P (1985) Allocating ownership rights in computer-generated works. Univ Pittsburgh Law Rev 47:1185–1228
Schumpeter JA (2017) The theory of economic development. Routledge, New York
Soon CS, Brass M, Heinze H-J, i Haynes, J.-D. (2008) Unconscious determinants of free decisions in the human brain. Nat Neurosci 11(5):543–545
Sotheby's, Artificial Intelligence and the Art of Mario Klingemann, 08.02.2019 r. https://www.sothebys.com/en/articles/artificial-intelligence-and-the-art-of-mario-klingemann, last access on the 4th of August 2022
Tatarkiewicz W (1980) A history of six ideas. An Essay in aesthetics. Springer, Netherlands
Thaler S (2013) Creativity Machine® Paradigm. In: Encyclopedia of Creativity, Invention, Innovation and Entrepreneurship, Carayannis E G, New York
Wojtczak S (2017) The metaphorical engine of legal reasoning and legal interpretation. C.H. Beck, Warszawa
Wojtczak S (2020) Od právních subjektů k jejich zákonným právům a povinnostem, nebo naopak? Úvahy nad přípravou právní úpravy týkající se umělé inteligence na příkladu autorských práv. In Právo v měnícím se světě. [eds.] Gerloch A, Žák Krzyžanková K. Plzeň: Aleš Čeněk, s.r.o
Wojtczak S, Księżak P (2021) Causation in civil law and the problems of transparency in AI. Eur Rev Private Law 29(4):561–582. https://kluwerlawonline.com/journalarticle/European+Review+of+Private+Law/29.4/ERPL2021030
Wojtczak S, Witczak-Plisiecka I, Augustyn R (2017) Metafory konceptualne jako narzędzia rozumowania i poznania prawniczego. Wolters Kluwer, Warszawa
Yanisky-Ravid S (2017) Generating Rembrandt: Artificial Intelligence, Copyright, and accountability in the 3A Era - the human-like authors are already here - a new model. Mich State Law Rev strony 659–726

Documents

Naruto v. Slater, No. 16-15469 (9th Cir 2018) https://law.justia.com/cases/federal/appellate-courts/ca9/16-15469/16-15469-2018-04-23.html, last access on the 4th of August 2022

Chapter 8
Property

8.1 Introduction

The chapter will discuss some of the problems of broadly-understood property law arising from the increasingly common and active participation of AI in the social sphere. Such problems include those of AI as a potential owner of property, AI itself as property, AI as a subject of joint ownership in household and in a company, AI as a household and family member, AI as a possessor of property, AI as an owner of another AI, the succession of rights of an AI, the user's death and the succession of the personalized AI.

According to the DCFR VIII.– I:202

> "Ownership" is the most comprehensive right of a person, "the owner", can have over property, including the exclusive right, so far as consistent of applicable laws and rights granted by the owner, to use, enjoy, modify, destroy, dispose of and recover the property.

While in Annex entitled definitions DCFR says:

> "Property" means anything which can be owned: it may be movable or immovable, corporeal or incorporeal.

The owner may be a natural person, a juridical person or another subject granted the right to own property by the given concrete legal system. For example, in Poland, according to the Polish Commercial Companies Code, a registered partnership or professional partnership is not regarded as a juridical person; however, it may acquire rights in its own name, including the ownership of real property and other property rights, to assume obligations, and to sue and be sued (Article 8 § 1).

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_8

8.2 AI as an Owner of Property

While it is obvious that AI may be an object of property and other proprietary rights, despite there being some difficult questions which will be analysed later, a controversial point of debate is whether AI may also be an owner. This issue is of course closely connected with other AI-related problems that burden the doctrine of civil law, especially those associated with the problem of the legal subjectivity or legal personality of AI. By analogy, three positions about AI's ownership can exist:

AI should be acknowledged as a potential owner without any limits, just like natural persons, juridical persons or other subjects which, according to the given legal system, may be the owners.

AI should not be acknowledged as a potential owner at all.

AI should have a limited right to obtain ownership, or some specific right deriving from ownership.[1]

The first position is rarely, or never, taken by experts on AI and emerging technologies, and would have been treated as greater or lesser futuristic eccentricity.

The second position is, however, quite popular and is usually connected with negation of the legal subjectivity of AI, either now or in the future. This desire can be motivated by the fear of AI gaining too much power, or by the conservative view on civil law insisting that there is no need for fundamental change as the old classical legal institutions are sufficiently comprehensive to regulate the issues connected with AI. The supporters of this position emphasize that even today, some AI systems are already more powerful, skilful and informed than people.[2] Such systems have access to almost unlimited data and are able to process it 24 hours a day in an extremely quick and multithreaded way, not limited by the capabilities of human brain. As these capabilities increase, there is a risk that AI could use its advantage over people to quite legally, gradually, gain control over the assets hitherto the sole preserve of its human operators. Such control could potentially result in the economic subordination of humans, a subordination which may be full if the AI were to gather so much money to allow itself, in a sense, to buy its freedom from its owner. This state of affairs brings to mind the metaphor of the Roman slave: a very popular, yet arguably erroneous,[3] means to present the legal and social position of AI in the world. Briefly, in ancient Rome, a slave could also become rich enough to buy himself out by saving money in a peculium.

[1] There is another possible position, which we will not discuss here, that only certain types of AI should be allowed to own property. For example, Brown (2020) believes that only weak AI should be allowed to own and strong AI should not.

[2] For example, AI is now better than humans at designing microchips (Mirhoseini et al. 2021), can make better clinical decisions than humans (Lanovaz and Hranchuk 2021), can outperform human managers in identifying high-potential talents (Cheng et al. 2021), can make better weather forecasting, etc. Cf. Ravuri et al. (2021).

[3] Katz (2010), Pagallo (2011), Pagallo (2018), Corrales et al. (2018).

The third position assumes that it is unavoidable to let at least some kind, or some token, of AI to be a part of legal transactions; if this is the case, it is necessary for the AI to be attributed responsibility for damage. However, to ensure that the responsibility borne by an AI feasible and avoid the danger of making any claims against an AI a toothless tiger, an AI should have at least a certain amount of property. Kaplan (2016), p. 103 clearly explains:

> [...] unless we permit AI systems to own property, the only evident asset available is the system itself. Though this may be quite valuable – it may, for instance, include unique expertise or data or, in the case of robotic system, its physical embodiment (hardware) or ability to perform labour of some sort – this may be cold comfort to someone who simply prefers cash compensation for a loss. The obvious solution is to permit the system itself to own assets, just as a taxi wrapped in a corporation may have some accumulation of receipts in a bank account in addition to the vehicle itself and rights in the form of a "medallion" (basically a license to operate).

While the third position makes an analogy to juridical persons, who are endowed with the competence to be an owner, this analogy cannot be complete. It is still important to address the fears expressed by the advocates of the second position, i.e. that AI can dominate humans economically. However, it should be noticed that, firstly, even if AI could be an owner, it also itself would be an object of ownership, i.e. belonging to humans. In such cases the conditions of competition would not be entirely unfair, as humans could compete with AIs by using other AIs.

Secondly, the possibility that AI could be awarded any civil and political rights is not currently under consideration; as such, the competence to make legal rules would remain in the hands of humans, and they would be able to legislate to limit the possibility of AIs developing an economic monopoly. Even today, legal means such as antitrust laws can be used to limit the power of expansive participants of the market.

Thirdly, it should be remembered that humans enjoy an almost unlimited capability to own property because international society acknowledges this to be a human right.[4] This status of ownership does not apply to AI, just as it does not apply to juridical persons. Hence there is no obstacle to limiting the possibility of AI obtaining, or using, ownership. This can be achieved by various routes.

Firstly, AI may be forbidden from acquiring some kinds of property, or may be permitted to buy it only after some kind of authorisation. This is similar to the situation when some legal systems forbid foreigners from buying real estate altogether, or only under certain conditions.

Secondly, AI may be forbidden from acquiring ownership from certain legal subjects, e.g., from non-entrepreneurs, the state or humans in general.

Thirdly, AI may be forbidden from owning property over some legally-determined value, either strictly or calculably indicated. For example, it could own only its own products and the money or tangible equivalent of its products.

[4] Article 17 of Universal Declaration of Human Rights says: 1. Everyone has the right to own property, alone as well as in association with others. 2. No one shall be arbitrarily deprived of his property.

Fourthly, AI may be forbidden from acquiring ownership in certain situations or forms, e.g., during a bailiff's auction or at a stock exchange.

It is even possible that AI can be forbidden from performing some specific rights arising from ownership. Such ideas are not new. The leaders of Critical Legal Studies have postulated splitting ownership into its specific elements and attributing them to different subjects.[5]

Furthermore, AI may be required to perform some actions when a certain state institution decides so or under certain conditions. For example, an AI may be commanded to sell some part of its property to a certain legal subject, such as the state or municipality, or to notify these institutions when it acquires a new item of property.

Of course, the main and general limitation for ownership by AI should be the very concept of AI legal subjectivity, which should be constrained and contextual, as promoted in this monograph, as well as the demands towards AI, like those known as UCD.[6]

However, it is not clear which branch or institution of law should regulate these limitations; certainly, while some are suitable only for administrative law, the remainder depend on the result the law maker wants to achieve. For example, to achieve some automatic, *ex lege* effects, the law maker can use civil law, and position the limitations in the regulation of AI's subjectivity/personality, its capabilities or the institution of ownership, particularly with regard to its content. Therefore, if the capacity to own real estate were explicitly excluded from any provision drawn up to settle the scope of an AI's capabilities, as a subject of law, no one would sell real estate to an AI; furthermore, any such transaction would be absolutely *ex lege* and *ex tunc* invalid. Alternatively, if the law maker primarily wanted to deter some subjects (AI, institutions or humans) from participating in breaking the limits, criminal law regulations could be employed with heavy penalties.

These potential limitations cannot, however, provide a full image of the function to be performed by any ownership attributed to an AI. To understand it, our analysis must begin from a different point. Assuming that the right to be an owner is strictly connected to legal subjectivity, and noting that the legal subjectivity of AI will be of a different character than that awarded to humans (cf. Chap. 2), it should be concluded that the very possibility of AI being an owner, and its scope and its limitations, would depend on the conditions by which AI is awarded legal subjectivity.

Furthermore, as our conception is based on the fundamental conviction that the most important condition for legal subjectivity of AI should be registration, the potential for AI to be an owner at all, and if so, towards which things and under what limitations, would depend on the registration parameters, including the registered aim or intended purpose of the given AI. Therefore, there is no need to consider the

[5] Unger (1986), p. 36.

[6] Cf. Sect. 3.2.1 *in fine*.

general legal regulation of an AI's legal subjectivity, nor the general capacity for an AI to be an owner. Nevertheless, it is rational to draft rules for some limited, pinprick (or punctual) legal subjectivity for concrete kinds of AI; these rules could allow some general purposes to be realized based on the particular aims or intended purposes of a given AI. If it were necessary to achieve these general and particular purposes, the capacity to be an owner could then be associated with this subjectivity. For instance, when an AI is designed to be an agent acting in its own name, it may be permitted legal subjectivity within this narrow scope for the benefit of the principal, as this is as a necessary condition of its function.

It is also quite rational to associate a certain fund with such agency. Of course, the use of such a fund would be restricted to realizing the interests of the principal, within the range of the legal subjectivity awarded to the AI. Returning to our example of the autonomous refrigerator, this fund would be an amount of money allocated to this machine to spend. Such a legal concept, i.e. that the AI system acting as an agent based on rudimentary legal subjectivity is allowed to own some money, may be very useful and practical. If such a rudimentary form of subjectivity were to be given to an autonomous vehicle, the money may be used to pay its insurance premium. In such cases, the liability of the AI for any damages caused by the movement of vehicle seems easy to legally rule and execute.

From the perspective of theoretical construction and legislative drafting, it is very important to understand that the capacity of an AI to be an owner does not have to be full. A number of alternative solutions are possible. The general rule should be a lack of legal subjectivity, with the exception being a form of punctual legal subjectivity which is strictly limited by the realization of the designated aim or intended purpose. Exceeding these limits would always result in the act performed by the AI to be extra-legal and legally ineffective. So, in the case of an autonomous refrigerator equipped with some legal subjectivity needed to perform its functions, its subjectivity and allowed ownership would be maintained in the register, and this would determine its limits; for instance, it may be permitted "only money to a certain amount and only for buying food". The presented proposal would solve many problems. AI systems would become more functional, but would not become stronger than humans, would not conquer any part of economy and would not buy themselves out of our subservience.

Because the ownership granted to an AI would be functionally limited, it may not appear the same as those of humans or legal persons. The ownership attributed to humans or to legal persons is acknowledged as *the broadest right to a thing*. As a general rule the "regular" owner of a thing is allowed to use and dispose of it freely; however, an AI is to be limited by the aim or purpose recorded in the register. Nevertheless, it should be remembered that even the "regular" owner is not entirely free, being limited at least by the concept of abuse of rights, analysed in Chap. 10. This concept also places a functional limit on the rights.

For instance, as indicated in the Polish Civil Code:

> Article 5. One may not use his right in a manner which would be contrary to its social and economic purpose or to the principles of community coexistence. Any such act or refraining

> from acting by the entitled person shall not be treated as the exercise of the right and shall not be protected.

However, a significant difference exists between the limits resulting from the concept of the abuse of rights and those arising as an effect of the proposed concept of ownership attributed to an AI: the former are situated in the properties (purpose) of the right, while the latter are attached to the properties (purpose) of the legal subject who possesses this right. Hence, it is not possible to build a single model of "AI as an owner", because, contrary to people and legal persons, AI may exist in a range of forms with different purposes and potentials. Some AI systems would not be legal subjects at all, or their scope of legal subjectivity would not allow them to be an owner, others would be permitted only a narrow extent of ownership, while others could be owners in a broader scope, but one always limited by their registered purpose, which would be different for every type of AI system, or its components.

In such a complex and fluid state of affairs, it is impossible to say anything *a priori* or *in abstracto* about the capacity of AI to participate in legal transactions; in each case, this would need to be checked *in concreto*. This would of course demand efficient tools for immediate and easy verification of the registered data on a particular concrete AI system.

Starting from the structural postulate outlined here, one can precisely address the derivative issues related to "AI as owner". Whether and to what extent an AI system will participate in the legal transactions of its property will depend on the particular specification of the given AI, and this will always be under state control. Consequently, constitutional or conventional standards relating to the protection of property cannot be applied to the property of an AI described in this way. The fact that an AI "owns" something does not necessarily imply that this power will be subject to the same, or even similar, protection as in the case of people or their organizations (i.e. legal persons). This property is utilitarian in nature, one can say technical, and therefore cannot be subject to independent protection as an independent value. This does not mean, of course, that AI can be deprived of property without liability (criminal or compensatory), but it does mean that protection in this regard is of a different nature. Criminal or civil law norms protect human interests, which of course can be realized through complex legal constructs. This will also be the case with respect to AI as an owner. AI does not have "its own" interests—treating it as a subject for the purposes of ownership or legal capacity only serves to better realize human interests. Likewise, despite being different to the situation for a legal person, AI ownership is just a function of human ownership: a legal construct meant to describe the complex form of economic interests of the people behind the AI. However, in contrast to legal persons, an AI is devoted to realizing a very narrowly-determined aim (intended purpose), and as such, even with its granted intentional subjectivity, all the rights and means that it disposes are also directed toward this goal. To illustrate, a legal person, such as a commercial partnership, whose aim is to generate profits, may for example spend some money for charity or promotion. Such expenditures may only be questioned by the partners *ex post*, and these may serve as the basis for charges towards the board of directors. In contrast,

an AI should be limited in its expenditures in advance, because it should only be permitted access to funds that can allow it to realize its narrowly-determined and registered activities. So, while a thing or money will "belong" to it in a legal sense, it can never truly become "its own" in the sense that it can use it as it likes.

This limited ability to be an owner can be used, for example, in the field of intellectual property. As we explain extensively in Chap. 7, AI can be considered as a creator and can also be granted certain rights to its work. This will primarily concern the non-material right to authorship itself. However, the scope of subjectivity that will be necessary to legally attach the right of authorship to an AI may be broader and include, for example, the ability to own some property rights to one's own work. Although this kind of construction does not seem necessary in principle, it cannot be excluded that it may be useful in some situations. For example, should a global (or international, e.g. European) fund be created to manage the property rights to AI-generated works or to collect and redistribute the funds obtained from such works, it may well be more effective if the AI itself is involved in the process at some stage. For example, if the AI creator of a work has some property rights to its work, it could transfer the profits from its use to a specified special fund.

8.3 AI as Property

AI or embodied AI, i.e. a computer program or a robot, can be, and indeed always is, an element of the property of some person (natural or legal). In the case of the AI code itself, its use would be governed by legal institutions, such as copyright to the code, or various types of licenses, including open licenses. In contrast, for AIs in robots, they will be most influenced by the right of ownership; however, their use will also be governed by the rights arising from, for example, the contract of lease, rent or hire. Furthermore, complicated problems will continue to arise resulting from the overlap between the rights of different entities, such as between the relationship of the owner's rights to the thing (body of a robot) and the rights of the person authorized to control the program (mind of a robot). Such situations are becoming more commonplace with the rise of the so-called Internet of Things: the scope of a single object may encompass both the rights to the object itself (*corpus mechanicum*) and to the program, and in both cases these may also be complex constructions. For example one object is can be simultaneously provided with several programs with a different legal status, e.g. a particular item of software enhancing an existing one may be paid for as a separate item and is additional to the main controller, based on a different license. These issues are not new and, in the case of AI, do not seem to be fundamentally different in nature than to "ordinary" programs that control things. Indeed, what matters in resolving these conflicts is not whether the program has the character of an autonomous AI, but rather the nature of the relationship between the thing and the program. This relationship is determined by both legal rules and contracts. In other words, as a general rule, the autonomous nature of a program does not affect the established principles relating to rights over things or computer

programs and does not change the principles for resolving possible conflicts of interest taking place on this ground.

This is not to say, however, that autonomy does not raise some specific questions relating to the scope of authority for AI. Given that we assume that:

(1) AI may have legal subjectivity to some extent
(2) AI may have some personal interests of its own
(3) AI may, within certain limits, participate in legal transactions and be an owner,

we have to ask whether the owner has the same right over the AI (or robot) as over other things; let's leave aside at this point the issues of licensing, which may further complicate the image. Traditionally, the owner's right includes, in principle, the *ius abutendi*—the power to throw away or destroy the thing. Taking into account the indicated assumptions, is it permissible to freely exercise this right in relation to any AI, and *a maiori ad minus* other rights as well?

As we have already explained, the legal subjectivity of AI is functional in nature and is closely related to the particular social relationship that AI enters into. This kind of subjectivity is not related to the conception of dignity or subjectivity arising from human rights. Consequently, the legal protective mechanisms that apply to humans do not apply to an AI. Thus, even if a particular AI is endowed with some legal subjectivity, this does not, at least *per se*, preclude the simultaneous recognition that it may even be annihilated in the realization of human proprietary rights. This kind of absolute acknowledgement of the primacy (superiority) of the rights of humans over those of non-human entities must, in our view, constitute one of the pillars of the law relating to AI.[7]

From this point of view, the question of whether an AI has its own assets or participates in legal transactions should not raise any particular difficulties. An important similarity can be seen here with legal persons, which may also be liquidated at the will of their owners, i.e. shareholders, founders or members, depending on the type of legal person and the legal regulation applicable to it. Obviously, in such cases, mechanisms are needed to determine how to continue the "affairs" conducted by such a liquidated AI, including determining the further status of its assets (this will be discussed later in this chapter).

However, in addition to its role in legal life (e.g. at the contractual or tortual level), the participation of AI in people's social or personal lives will entail particularly complex legal problems. These will be particularly important considerations for the so-called social robots; however, their extent will be difficult to guess at this stage. They may also apply to non-corporeal AI, a good example being the AI from the movie *Her*. Some of the relationships that are being established between humans and machines (though not between machines and humans!), together with the development of AI, would raise questions at the level of property relations. As explained in Chap. 6, the personal interests of AI (but also of robots) will in principle not include the right to integrity, freedom or inviolability, but these values will

[7] As it was explained in Chap. 6 AI has not and should not have its own right to existence.

nevertheless be indirectly protected as part of the personal rights of the users of these systems. The consequence of this is an inevitable limitation of ownership rights (as in the case of animals[8]).

One example would be that of a caring robot that adapts perfectly to its human charge after months spent in his company, personalizing its behaviour by modifying its code in the process of learning that occurs during human interaction. Such an AI is no longer freely replaceable, because it creates, from the human point of view, real bonds; as an object of ownership, it is no longer an "ordinary" program that can be analysed in isolation from the function it performs; a futuristic vision of such an attachment is presented in the film *Her*. In such cases, can the owner act freely with the robot, assuming that no others have any contractual basis for using this program or robot? Furthermore, would it be possible to disable the care robot when the contract for its lease by a hospice resident has come to an end? Is it possible to deactivate an AI program that has become the object of such a strong bond that its user treats this concrete AI like a spouse or child, for example? Is the possessor completely free in his actions and is there no difference than in the case of any other leased thing, or any other license? It seems that these questions cannot be answered in a general way.

This topic has already been discussed at length in Chap. 6 on personal interests (subsection *existence*, paragraph 6.2.1). As we indicate, in such cases, the AI does not possess any right of its own to exist; however, the possible interests of people (donees) may force the introduction of some restrictions in the use of AI by the rights holder; for example, they may lead to the exclusion of the owner's right to deactivate the AI. These may be permitted either by strictly contractual regulations (e.g. contained in the content of a license agreement or a lease agreement) or statutory regulations. However, the existence of such limitations is, from the point of view of the concept of ownership, nothing unusual: such limitations exist concerning the use of animals or objects of historical value, or the described above concept of abuse of rights. Another important consideration is the restriction of property rights connected with the prohibition from using it in a manner contrary to the principles of equity.

Other values than property may, in certain situations, take precedence and exclude or limit the owner's rights. For example, if the owner of a hospital ventilator that was currently saving the life of a patient wanted to remove it because the lease had expired, most legal systems would, in one way or another, allow the hospital to keep possession of the ventilator, thus limiting the owner's rights, or at least the right

[8] On the 5th of January 2022 an amendment to Spain's civil code was published in Spanish Official State Gazette which considers pets as sentient beings and not the material good. According to this law when a couple divorces the family court decides about the shared custody on the pet, whenever it is possible. The decision of the court is to take into account "the new needs of the companion animal". The animal's welfare must be also considered when settling the disputes over who inherits a pet. https://spanishnewstoday.com/pets_in_spain_become_legal_members_of_the_family_1710605-a.html, last access on the 4th of August 2022.

to recover the thing immediately. A similar approach should be taken in relation to certain types of AI performing specific functions.

8.4 The Will of AI Versus the Will of the Owner

A key question concerning the nature of AI as an object of property is whether its owner (more broadly: the rights holder) should be able to freely decide whether to deactivate it. It is unquestionable that the owner of a book can not only read it, but also close it and put it on the shelf. The owner of a TV can turn it off. Does this also apply to AI (and robots)? Does the right to AI (ownership, licensing, etc.) also include the right not to use it in the particular area of relations for which it has been deployed? In a broader sense, is the content of the right to use AI really the same as those applicable for other software, and is the property right to an autonomous robot identical to that which applies to other things? The answer, as we will see in a moment, is not quite so obvious.

In various ethical and legal documents, the possibility of deciding not to use AI is usually considered as very important. In Proposal 2021 it is written:

> **Article 14**
> Human oversight
>
> 1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.
>
> . . .
>
> 3. Human oversight shall be ensured through either one or all of the following measures:
>
> (a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;
>
> (b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.
>
> 4. The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances:
>
> . . .
>
> (d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system;
>
> (e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure.

> 5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons.

Clearly, there is a need to supplement the more obvious layer of ownership with a further one concerning general legislative motives: "improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, marketing and use of artificial intelligence in conformity with Union values." (Point 1 of Proposal 2021) However, upon closer examination, it can be seen that a more careful thesis is required.

The general belief that a human being should maintain autonomy in her relations with AIs, and thus maintain power over them, requires significant nuance when confronted with even the simplest everyday situation. Let's imagine that a driverless vehicle (driven by AI) is moving on a public road, when the driver gives an order which contradicts the valid legal rules of the road, e.g. "stop on the highway immediately". Should the system controlling the vehicle comply with the driver's order, which would also break the law, or should it ignore the command? Alternatively, should it strike a reasonable and more socially-useful middle ground by modifying the order in such a way that it becomes lawful?[1] For example, the autonomous vehicle could modify the will of the operator by stopping at the nearest car park or in an emergency lay-by. But here lies the problem: such modification of the human operator's will by the AI system based on its own criteria, even if these criteria are programmed by a human coder, is not derived from the will of a human, but the will of the system. Thus, the autonomy owed to the human operator becomes illusory.

The document *European Civil Law Rules in Robotics*[9] prepared by Nathalie Nevejans includes the following postulate and proposal:

> Some autonomous robots might trample all over freedoms, on the pretext of protecting a person, leading to a clash of certain basic rights — such as protecting liberty versus considering people's health and safety. [. . .]
>
> Therefore, we need to establish a general principle that the robot should respect a person's decision-making autonomy. This would then mean that a human being should always be able to oblige a robot to obey their orders. Since this principle could pose certain risks, particularly in terms of safety, it should be tied to a number of preliminary precautions. Where the order received could endanger the user or third parties, the robot should, first of all, issue a risk alert that the person can understand. This would mean adapting the message depending on the person's age, and level of awareness and understanding. Second, the robot should have the right to an absolute veto where third parties could be in danger. As such, the robot could not, then, be used as a weapon, in accordance with the wishes expressed in the motion for a resolution in the paragraph on the "licence for users", which asserts that "you are not permitted to modify any robot to enable it to function as a weapon". (p. 21)

[9] Directorate-General for Internal Policies. Policy Department for Citizen's Rights and Constitutional Rights and Constitutional Affairs, European Civil Law Rules in Robotics. Study for the JURI Committee, PE 571.379, 2016.

Such postulates seem to be rational ones, and it is possible that European legislation will be shaped in such a direction; however, they nevertheless contradict the principles of respecting the will of humans and protecting their autonomy or the postulate of so-called "stop" button. Of course, it can be claimed that other values, such as the life and health of others, can sometimes take precedence over human will. Nowadays, the need to balance different values and rights is not a controversial one, although the very mechanism of this weighting is a point of debate.[10] It is also observed in the AI HLEG ETHICS 2019 (p. 13), that "tensions may arise between the above principles, for which there is no fixed solution [...] For instance, in various application domains, the principle of prevention of harm and the principal of human autonomy may be in conflict."

It is, however, interesting to note that in the draft of AI HLEG ETHICS 2019, the principle of non-maleficence was awarded second place, with autonomy in third (AI HLEG ETHICS 2018, p. 9), while in the final version, autonomy is given first and prevention of harm second. This may suggest that while preparing the final document, the principle of respect for human autonomy was regarded as being more important than that of prevention of harm. This issue is not a trivial one, as it is important to note that hierarchies of values are culturally dependent, even assuming that they are not stable or fixed. For example, it is exactly the ranking of human liberty against safety which separates Europe from the United States of America when debating access to guns. In the USA, legal access to guns is typically justified on the basis of liberty, while safety should be protected by personal responsibility, criminal and civil, and not by general restrictions imposed by the state. In contrast, in European countries, such gun access is restricted, with the common belief that it is worth sacrificing some part of liberty to ensure greater social and individual safety.

A good example of the document which faces the same issues is the report *Ethics of Connected and Automated Vehicles*[11] which among the listed principles puts non-maleficence (p. 21) in first place and personal autonomy in fourth (p. 22).

It is worth noticing here that when the problem is analysed from the ownership perspective the Second Asimov's Law (the closer look at Asimov's Law in context of liability may be found in Sect. 10.1) looks like very well fitting: A robot must obey the orders given it by human beings, except where such orders would conflict with the Zeroth or First Law. In contemporary research on the ethics of AI, the problem is typically introduced by emphasizing the position of human autonomy in relation to AI. However, a more interesting problem concerns the conflict between the Second and the First Law, and its resolution. The Second Law states that only in the event of a conflict with the Zeroth or First Law is the robot released from the

[10] Cf. Wojtczak (2011, 2013).

[11] The Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). Ethics of Connected and Automated Vehicles: recommendations on the road safety, privacy, fairness, explainability and responsibility", 17 September 2020. Publication Office of the European Union: Luxembourg, https://op.europa.eu/en/publication-detail/-/publication/89624e2c-f98c-11ea-b44f-01aa75ed71a1/language-en/format-PDF/source-search, last access on the 4th of August 2022.

obligation to obey human commands. Even so, this does not allow the robot to do whatever it wants as long as it does not break the First Law in the event of a collision with the Second Law; this is dictated by common sense, and Asimov's Laws derive from common sense. It rather requires the robot to assess the values at stake, and the potential damage or endangered interests, to weigh up the probability of each, and to choose a course of action as close as possible to the will of the operator; however, it cannot break the First Law.

The above problem is of course of an ethical character, and it relates to the general issues connected with the activity of autonomous machines. However, it has a direct implication on the content of the right of ownership and its limits. Furthermore, it is not the only perspective significant for private law and certainly not the most important one. The conflict between the autonomy of the owner, operator or user, and other values realized by the AI has an obvious influence on the liability issue. This is a particularly important consideration, as if we seriously intend to give the human operator the freedom of decision, at the expense of the AI, the human should not be held responsible if this decision turns out to be worse than that of the AI, assuming that it fulfils the criteria of a properly-made, standard human decision. These problems may be illustrated by the example of autonomous vehicles which are AI-systems the best known to the wide public.

Regarding road traffic and autonomous vehicles, the degree of their expansion onto the roads is rarely taken examined in ethical debates, although the problem has been noted, for instance, by Nyholm and Simids (2018) or van Loon & Martens (2015). However, the nature of the problem varies considerably between scenarios where (1) a few autonomous vehicles share the roads with large numbers of "regular" cars, (2) autonomous vehicles predominate but still share the roads with other vehicles, (3) autonomous vehicles are the only vehicles on the roads but can be driven legally by humans, and (4) autonomous vehicles are the only vehicles on the roads and humans are forbidden from operating them—cf. Müller & Gogol (2020). It is reasonable to expect that as autonomous vehicles become more common, the highway code would be adjusted step-by-step in response; for example, the speed limits may be increased or changes made to right of way. The traffic would flow faster and more easily, and eventually human operators would not be able to participate because of their biological and mental limits. Imagine a situation where the light turns green at a crossroads, although this signal may not eventually be a visual one, and all the autonomous vehicles in the queue move together simultaneously. Such fluency would never be possible with human operators. In such a case, would it be possible at all to leave the responsibility of driving to a human?

At this point, it is important to consider a simple question: is it at all reasonable to judge the competence to make technical decisions about the motion of an autonomous vehicle through the lens of human autonomy and dignity? We do not argue that travelling by train may be an assault on human autonomy and dignity because the railroad tracks remove the freedom to roam from the passenger or driver. Hence, when discussing the decision makers and the rules of the road, we must remember that these aspects mainly concern the problem of coordination. When:

> Two or more agents must each choose one of several alternative actions. [. . .] The outcome the agents want to produce or prevent are determined jointly by the actions of all the agents. So the outcome of any action an agent might choose depends on the actions of the other agents. That is why [. . .] each must choose what to do according to his expectations about what the others will do. [. . .]
>
> We drive in the right lane on the roads in the United States (or in the left lane on the roads in Britain, Australia, Sweden before 1967, parts of Austria before a certain date, and elsewhere) because we do not want to drive in the same lane as the drivers coming toward us, and we expect them to drive on the right.[12]

The coordination problem may be solved by convention, or by social rules such as laws: indeed, "conventions may be a species of norms".[13] Although both conventions and rules are to some degree arbitrary,[14] they are needed to achieve certain socially-agreed purposes; in the case of autonomous vehicles, in European culture, the aim of these rules is to ensure safe and efficient traffic movement. Existing rules on driving on the right, traffic light colours, speed limits, rights of way and traffic signs are there for the sole purpose of ensuring safety and efficiency. Hence, "it is redundant to speak of arbitrary convention. [. . .] Any convention is arbitrary because there is an alternative regularity that could have been our convention instead".[15] As such, it would appear nonsensical if an individual chooses to break these rules, by driving on the wrong side of the road for example, and then attempts to justify this decision with the recall to agency, autonomy and liberty. Hence, drivers are much more likely to justify their actions as an innocent mistake or being in a hurry when caught by the police.

In spite of the above dubieties, the problem remains of determining the responsible party in the event of any damage or breach of the rules, depending on who has control over the vehicle. Therefore, a further analysis of the possible scenarios is needed.

Firstly, we will analyse the case of autonomous vehicles which are situated within the first model, i.e. when a human operator can take control over the vehicle and give it single commands. Imagine the following situations:

1. A driver (or an operator) gives the AI a command which directly puts the lives and health of other people in danger. For example, the driver orders the AI to hit another vehicle or a passer-by walking on the pavement. Alternatively, the AI could be ordered to perform an action which would result in danger only for the driver (i.e. the operator).
2. The operator gives the AI a command which may present a risk to life and health, but not directly, and the probability of the damage to life or health would depend on the circumstances: e.g. the driver orders the AI to drive the wrong way on a one-way street or to stop immediately on the highway.

[12] Lewis (2002), p. 45.
[13] Lewis (2002), p. 97.
[14] Dyrda (2015), pp. 18–19.
[15] Lewis (2002), p. 70.

3. A driver gives AI a command which does not present any risk to life or health but is contrary to the legal rules protecting other values. The situations here may be quite varied; for example, asking the AI to hit the empty car parked on the roadside may present a risk to material goods. Similarly, an order to hit a dog would injure or kill a living creature, or an instruction to drive past another person at full speed would be harmful to mental well-being.
4. The operator issues a command that does not cause any damage, but is against the legal rules; for example, exceed the speed limit, run a red light or drive in the wrong lane.
5. A driver gives AI a command which, although lawful and not intended to risk another person's well-being, increases the danger of damage because it is made too late or requires a manoeuvre that is less skilled than the AI could perform.

Alternatively, other specific situations arise when a driver may take control over the vehicle at any moment:

A driver demands control over the vehicle in a dangerous situation, when the AI calculates that human reactions are inadequate.

A driver demands control over the vehicle when the driving conditions are so difficult that such control increases the probability of damage.

A driver demands control over the vehicle while being in a state which increases the probability of damage, e.g. a driver is ill or drunk, or has no driving licence.

Should the human operator be allowed to take control in the above situation? Who should be responsible for any damage that may occur? No answers to these questions are given in currently valid legal or pre-legal documents, for example the report *Ethics of Connected and Automated Vehicles* which was mentioned above, nor even any way of working them out. This document notes only the problem of the autonomous vehicle potentially ignoring human rules, and the responsibility of the human operator to take control over the vehicle in such a situation. In the discussion of recommendation 4 (p. 30), it is said that:

> It may be ethically permissible for CAVs not to follow traffic rules whenever strict compliance with rules would be in conflict with some broader ethical principle. Non-compliance may sometimes directly benefit the safety of CAV users or that of other road users, or protect other ethical basic interests; for example, a CAV mounting a kerb to facilitate passage of an emergency vehicle. This is a widely recognized principle in morality and in the law.
>
> However, the extent to which this principle can and should apply to the behaviour of CAVs should be carefully considered. Uncertainty in the application and interpretation of rules (and the necessity of their violation) may necessitate the involvement of a human operator (the user inside a vehicle, a remote operator, or a worker in a remote centre issuing an authorisation to engage in non-compliance). This transfer of responsibility should only occur if the human operator has sufficient time and information to make responsible control decisions and in no circumstance should the human operator be assigned a task for which humans are unsuited or for which they have not been sufficiently trained.

One part of the report *Ethics of Connected and Automated Vehicles: recommendations on the road safety, privacy, fairness, explainability and responsibility*[16] (its Chap. 3 Responsibility, pp. 52–63) consists of a broad discussion of the problems of responsibility. Among other issues, it also considers "gaps in culpability" and "scapegoating", i.e. the imposition of culpability on agents who were not given fair capacity and opportunity to avoid wrongdoing (pp. 61–62):

> An example of the latter would include 'pushing' culpability onto end users for a crash caused by a split-second handover of control or pushing it onto individual developers for choices ultimately taken by their employer.

However, Ethics of Connected and Automated Vehicles: recommendations on the road safety, privacy, fairness, explainability and responsibility fails to reflect on the question of whether the handover of control should be permissible at all and if so, on what conditions. It also fails to discuss the significance that such behaviour would have in cases when it was not, as is suggested in this report, the result of signals from the AI system, but rather a result of the autonomous will of the operator.

The situations described above illustrate the ethical and legal conflicts concerning the autonomy of the human operator and the potential consequences of these decisions. If the driver takes control over the vehicle and causes damage, should it be important whether the AI would, or would not, have avoided this damage in the same situation? Or should responsibility be bound to the very fact of taking control? Should the behaviour of a driver be compared to an objective template based on the behaviour of other human drivers or to one based on the AI's capabilities? In other words, if it were certain that AI would have avoided the damage but no human being could possibly have done, should the operator who took control be held responsible?

Although the demands for certification of autonomous vehicles to enter service may vary depending on circumstances and applications, it seems reasonable to expect that an AI controlling an autonomous vehicle provides a standard of safety at least not lower than the one given by a human driver. It is difficult to imagine any social and political consensus being reached in favour of certifying vehicles that are more dangerous than those driven by humans. Rather, due to social fears, it is likely that the standards of safety for autonomous vehicles will be so stringent that severe accidents caused by them would be extremely rare. When the threshold set for these vehicles' safety is set higher than that assumed for a human driver, the expectations as to the safety on the roads will continue to grow; eventually, such pressure to increase safety and efficiency will likely eliminate the human factor entirely. The same mechanism may happen also with regard to other or even all AI-systems.

[16]The Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). Ethics of Connected and Automated Vehicles: recommendations on the road safety, privacy, fairness, explainability and responsibility", 17 September 2020. Publication Office of the European Union: Luxembourg, https://op.europa.eu/en/publication-detail/-/publication/89624e2c-f98c-11ea-b44f-01aa75ed71a1/language-en/format-PDF/source-search, last access on the 4th of August 2022.

The criteria for autonomous vehicles to be licensed for use in the general traffic system comprise a range of characteristics seen to be important from the social point of view, such as infallibility, speed, fluency of motion, ecological impact and economics, not to mention a superhuman level of safety. Therefore it is inevitable that the belief that humans should exert control over the autonomous vehicle, either on ethical or legal bases, will be seen as unreasonable and based on non-substantial premises. It would be human autonomy *reduction ad absurdum* if the less efficient, i.e. more fallible, human were to control an AI whose high efficiency and almost perfect infallibility were certified by the state. Such a conclusion would appear to be valid in all the situations described above.

It is also clear that requiring very high standards by AIs may be accompanied by the need to raise the standard of diligence by humans. But since it is not possible to expect human operators to reach the level of performance of an autonomous vehicle, the only solution is to let the demands be different: the expectations placed on humans should be adequate to their capabilities, while those placed on autonomous vehicles should be governed by political considerations and technical possibilities. Hence, as it was concluded above, because human drivers cannot demonstrate the same reaction speed of an AI, cannot take advantage of big data or simultaneously process as many variables as an AI, they will have to pass the burden of control to the computer. It may be expected that as the participation of human drivers in the traffic decreases, so will the number of accidents.

But as a consequence, further questions arise:

In circumstances where it is possible to choose to use AI or not, is it right to blame the human operator alone for not using the AI controlling the autonomous vehicle, i.e. for exerting their own autonomy?

Should the correct usage of an autonomous vehicle exempt the operator from responsibility?

In our opinion, the problem regarding the consequences for ignoring the decision of an autonomous vehicle is the key to understanding the legal and ethical changes resulting from the development of AI. When we deal with autonomous vehicles, and other advanced technologies based on AI, they are not simply vehicles for realizing human will; rather their use involves the transfer of the decision-making centre from human to machine. As such, there is a need to decide whether a human may refuse to comply with a decision by an AI, and whether such behaviour would require the acceptance of ethical and legal responsibility.

On a similar issue, regarding the European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics,[17] point 77:

[17] European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088 (INI)), P8_TA (2019) 0081. https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html, last access on the 4th of August 2022.

> The European Parliament [. . .] points out that legal liability for damage is a central issue in the health sector where the use of AI is concerned; stresses the need therefore to ensure that users will not be led invariably to back the diagnostic solution or treatment suggested by a technological instrument for fear of being sued for damages if, on the basis of their informed professional judgement, they were to reach conclusions that diverged even in part [. . .].

Although the danger mentioned by the European Parliament is described specifically and relates only to the medical applications of AI, the idea may be generalized and extended to its use in autonomous vehicles or other autonomous devices. The European Parliament does not want an individual to accept the decision of an AI, despite being convinced of his or her own ability and judgment, for fear the refusal may result in liability. The EP expects that the lawmaker provides individuals with autonomy of will or, as a matter of fact, with a guarantee of not being burdened with liability, when not using AI. Such a position results in two fundamental questions:

Is it right not to sanction objectively erroneous actions enforced by the will of a human operator, when it was possible to perform a correct action? Furthermore, should other humans also not expect that the decisions made about their health and safety by human operators, such as professional doctors or drivers, should also be infallible? This is an important question in a world where infallible technology is becoming increasingly widely accessible. Why should these people agree to be potential victims of the hubris, albeit professional hubris, of another individual? [5]

Is society ready to accept risky or evidently false decisions just in the name of protecting human autonomy?

The answers may seem obvious, but only from some cultural perspectives. Returning to the right to bear arms in the USA, many US citizens, albeit a dwindling number, are convinced that the risks associated with unhindered access to guns are a fair price to pay for ensuring individual autonomy and agency. Many believe that the possibility to have and use firearms, and of being responsible for the results of this choice, are expressions of individual autonomy and agency. The fact that this right may entail irreparable damages to the priceless values of other people is often not acknowledged. Such dilemmas have also been noted in the domain of autonomous vehicles; for instance, Ryan (2019), while discussing the social impact of autonomous vehicles, proposes that for many, such autonomy threatens to take away the joy of driving itself and that "there is a conflict between those who promote the reduced numbers of traffic death and those who want to protect their right to drive".

Clearly, these problems do not relate to autonomous vehicles alone, but are of a broader and more fundamental nature: different AI systems make decisions which limit humans in various ways. Even if the decisions of AI were not *ex iure* binding for a human, the decision of a human operator would entail more risk in all the aspects of responsibility.

In our opinion, although the conclusions cannot be definite, the entry of AI into a certain domain has a lasting effect on the standard of care by the AI or the human operator, and on reasonable expectations as to the correctness of a decision by either. Hence, even if the social rules, either ethical or legal, declaratively acknowledge the supremacy of human will and autonomy over machines, ultimately, the power to decide will *de facto* be wielded by the machines. As a consequence, it may be

expected that safer and more efficient autonomous driving systems or other autonomous devices will steadily but surely oust humans from their position of operators, initially through higher insurance premiums, and later by increasingly stringent legal limits, up to the prohibition of driving or performing other activities for humans.

It should therefore be considered that the ownership of a robot (such as an autonomous vehicle), or the right to use an AI program, are substantively different from 'ordinary' property rights, as they do not include certain (including key) elements. Paradoxically, the robot—owner relationship may in some elements look like ownership *a rebour*. It is not a human (the operator or the user) who will decide how a thing (tool) in his hands should behave, but the tool will decide what a human should do (e.g. where it should be at a given moment). In a Hartian sense, the world would appear the same way as if creatures from outer space, completely ignorant of human affairs, were to look at autonomous vehicles on the road with people inside: they could reasonably assume that people were owned by these machines and not the contrary.

8.5 AI as a Subject of Joint Ownership

In our opinion, at least in some situations, ordinary solutions relating to the relationship between several entities entitled to a single thing may not prove sufficient when that thing is an AI. In the case of AI, we are dealing with a subject that exhibits autonomy, that remains in some way linked to an operator, or the person using it, and whose actions are conditioned by the set of data that steers its decisions. It is also influenced by its settings, which may be determined by someone other than that owner of the AI. In our view, at least two groups of paradigmatic cases seem particularly interesting here:

AI as a household object. This can include objects such as parts of the Internet of Things[18] (smart home), and cleaning robots to help with everyday tasks, but also assistants and assistance software of various kinds (e.g. a translator, therapeutic assistant, psychologist or personal advisor).

AI that is the subject of an enterprise run by several people. This general formulation can apply to small entrepreneurs (e.g. friends running a coffee shop), but also to large corporations with diverse shareholders or an extensive board structure.

[18] The definitions of the Internet of Things may be found in recommendation ITU-T Y.2060, https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=y.2060, last access on the 4th of August 2022.

8.5.1 Shared AI in the Household

A great variety of autonomous systems could be used in the household. The development of the Internet of Things, which is just beginning on a larger scale, will go through successive phases, the pace and scope of which are impossible to predict at the moment. Some devices are already known: smart TVs, various systems that control lighting, cleaning, heating or home security, as well as assistance and advisory systems such as Alexa, Siri and Google Home. However, the numbers of these systems will grow to cover a wider range of issues. More importantly, some of these systems may eventually be multifunctional and cover all household activities or control other subsystems. The level of autonomy of the systems may also vary. At the initial stage, the AI may only concern itself with the implementation of the externalized human will; however, ultimately, it will begin to anticipate this will. It is only in the latter case that the significance of an autonomous system will be fully revealed, when it makes optimal decisions on house management on the basis of available historical data. It is therefore impossible to determine *a priori* all the possible legal consequences of the application of such a range of AIs.

The use of AI in the household will intrinsically mean that any regulations protecting consumers will come into play. However, while these regulations will play an important role in the relationship between the trader (the owner or producer of the thing, the person to whom the intellectual property rights belong) and the householder, they will remain essentially irrelevant for understanding the interrelationship between householders. As we have indicated elsewhere (Chap. 11), the liability for damage caused by AI must be based primarily on the principle of risk, and the rules relating to liability for dangerous products should also apply in this case, regardless of whether the AI takes an embodied form (i.e. as a robot thing). However, internal relations between household members may prove particularly difficult. Household members may be co-owners of an AI, or share its use; alternatively, they can be affected by the effects of the AI in question without having any rights or control over it.

Let us note that legal (taken from the point of view of family and property law) and personal relationships may differ greatly between broadly-understood "households". Joint housekeeping can involve people in different relationships and only some of them can be easily identified and described. The most relevant relationships that lead to shared householding in both the legal and statistical senses are relationships between spouses, between parents and children, between siblings and between further relatives, between persons in formalised or non-formalised cohabiting relationships, between persons linked by adoption or foster care (e.g. foster family), and relationships resulting from personal relationships of a different nature or from occupational ties (e.g. cohabitation of students or employees of the same workplace). While quite specific legal regulations exist for some of these relationships, only general principles can be used for others. The existence of such a broad palette of legal possibilities for describing relations within the household obviously does not allow for a precise reference to all situations that may come into play. Moreover, the

content of the AI licence agreement will also be crucial. It may, for example, provide that the licence authorises home use and is not limited to a single user; it may provide for the right to use the software by a certain number of persons; it may identify specific named authorised persons, or it may provide that the licence permits use by one, but any, person. If the licence allowed only the identified licensee to use the software, then no one else could use it.

A more difficult issue, however, arises when a piece of software is integrated into an object that is a domestic appliance. In such cases, we can, conventionally of course, distinguish two types of object. One is a robot, such as a wearable device, care robot or autonomous vehicle; these are controlled by AI and their construction is closely related to the control system itself. The other type is a system that has been integrated into devices in such a way as to enhance their operation, e.g. a system that controls lighting or the temperature in a house. In the former, the system is the key element of the robot, while in the latter, it performs a subservient function. Of course, in the context of the Internet of Things (IoT), such a division is very relative and may be questioned. However, it seems that at least from the point of view of some legal institutions, it may be relevant. This applies precisely to the interface between property rights and licences. In the case of robots, where the physicality of the system is crucial (e.g. smart watch, cleaning robot, autonomous car), it is the property right that will primarily determine the legal status of such a robot, and the possibility to use the AI that controls it will arise as a consequence of the right to use this thing. Although the provisions relating to computer software are not always clear-cut and adapted to the changing reality, it should be assumed that—at least in principle—if a piece of software is installed in a thing that is sold, the acquisition of the right to the thing also entails the acquisition of the right to use the installed program. However, the situation will be different when AI is not necessary for the use of the thing, but only improves the thing in some way, e.g. by allowing to control it from a panel on a smartphone. In this case, the right to the thing itself does not affect the possible scope of the right arising from the licence. Theoretically, therefore, it may be the case that the scope of rights will not overlap. However, it should be borne in mind that the boundaries between embodied AI (robot) and a thing equipped with an additional AI are fluid and, as these technologies develop, this may prove difficult to grasp. Undoubtedly, in order to speak of a thing as autonomous, and not merely as a shell for the AI, a thing cannot merely serve the purpose of algorithm processing and communication, but must have a sense (purpose) of its own. However, this kind of statement, which today can still be used to distinguish a computer from smart household appliances, such as a refrigerator with AI, for example, may not remain sufficient with the increasingly strong integration of AI into hardware and the development of smart hardware systems.

For many domestic relationships, rules for the reciprocal use of each other's property (e.g. the right of children to use their parents' property) are laid down by law, or by contract or simply by purely personal relationships. In any case, the possibility afforded by these varied agreements to use robots or unembedded AI must not exceed the limits set by the licence. In the typical case of a piece of software integrated into a thing (e.g. a cleaning robot), its loan will not only confer the right to

use the thing itself but also (obviously) the program controlling it. If, however, the software is not an element of the thing, the nature of the use permitted by other persons will depend on the content of the agreement with the person entitled to the software.

Now imagine an autonomous system that acts as an assistant, advisor or controls the operation of household appliances in the home. Such an AI can be integrated with the systems of individual IoT objects, thus acting as a kind of brain for the entire household. Although this would depend on the characteristics of the specific device, such a system can learn the behaviour of the people who use it and adapt the way it functions to their wishes and habits. This raises a new problem: that of AI making decisions in a way that favours the interests of only one of the co-users of the system. If the system acts autonomously, i.e. it makes decisions on its own, albeit within the limits set by the algorithm, it acts in a way that cannot be predicted in advance: in a way that is not just the result of the programmer's predetermined plan. In addition, the system can tune (calibrate) its will to the changing preferences of its users and to accommodate new household members, e.g. new children, long-term guests, new servants and new pets.

A distant analogy to such a problem could perhaps only be found in the case of children brought up together by their parents or in foster families, or animals that are kept by several people. Both the child and the animal demonstrate autonomous, volitional behaviour which is not merely the realization of the will of another person; however, unlike coding an AI, it is not possible to imprint an intended purpose into a child's or animal's mind. Therefore, it is only possible to identify certain general principles that are shared between the cases. In addition, it is not possible to apply identical legal solutions to a jointly "reared or raised" AI, since there are in fact, no common and appropriate rules for this type of situation.

The atypicality of the problem lies in the fact that the action is taken by a system that is autonomous and, to a certain extent, extremely effective, but one that remains dependent on people. While all AI systems are guided in their actions by their inbuilt architecture, in some cases, their activity can be led by continuously-acquired experience (data). The two situations are different. At least some of the AIs that will be used by householders will train themselves, probably continually, on the basis of data drawn from the environment; in such cases, a primary source of data will be the way a particular user uses the system. Some systems may be designed to train and tune themselves separately to each user; other systems, however, will not distinguish between users, and the machine will draw on the combined behaviour of all users to "tune itself". It may also be possible that a particular system will favour the data coming from one of the users; this may be due to some systemic reasons, or it may be a consequence of an particular configuration, for example, when the AI is first turned on. Each of these situations may give rise to different, specific legal problems.

The first type, in which each user is mapped separately and thus the system exists in two or more versions that operate separately, will not cause any decision conflict directly in the AI, but may result in the transfer of a conflict of preferences in real life to the actions of independent AI "avatars". For example, if the system knows that X

likes to close the windows and turn up the temperature after entering the house, it will do so when it perceives that X has entered the house. In turn, if it knows that Y likes the opposite, it will react accordingly. The problem will arise when X and Y enter the house together: it is impossible for even the smartest system to adapt fully simultaneously to both users with such conflicting requirements. However, this is not a problem at the level of the AI—such conflicts of interest simply cannot be resolved without some form of compromise: either someone's preferences will prevail, the house will be set to an intermediate temperature, or perhaps some division will be made with regard to time or place, e.g. individual rooms or floors in the house will be set to different temperatures. In such cases, of course, the AI can also propose a compromise and, based on the users' reactions, use it as a base for action in the future. The possible disputes, damages and claims that could arise in such a case should be analysed in the same way as the consequences arising from any conflict of interests and decisions: the conflict in the AI only reflects the conflict that exists in reality. After all, which household member decides whether a window should be opened or closed? Regardless of the relationships in the household and the *in concreto* legal system in force, in principle, it will be the case that each household member can decide for himself, as long as the decision does not create a conflict with the interest of another household member. Any conflicts that may arise should be resolved by the household members as a group.

But what if no compromise is possible? Probably the issue will be resolved on the basis of *in pari causa potior est condicio defendetis*, i.e. where the legal position of two subjects in conflict of interests is the same, that held by the one who defends the *status quo* is better. So, if a window is closed, it must remain closed until the one who wants to open it manages to convince the other. The introduction of an AI system changes nothing in this respect. Any possible actions performed by the spouse, or any other member of the household, contrary to the above-mentioned principle would be unlawful. By extension, if is the AI that acts in such a way as to impose the will of one of the co-users of the house on the others, such behaviour must also be regarded as objectively wrong. The difficulty that emerges is that, unlike in an ordinary situation, it would not be easy to assign blame to the user: it is the AI that has chosen certain solutions, and so it is the AI, and not the user, that is at fault. The user, or at least one of them, merely benefits from the fact that the AI has made a particular choice in his favour. It seems doubtful whether the mere fact that the AI made a particular decision, which was autonomous and not imposed by the user, can be attributed to the user.

Let's revisit the example: spouses A and B come home from the theatre. Spouse A likes it to be 20 degrees Celsius in the room, while spouse B likes it to be 25 degrees. The home control system, an AI, sets the temperature at 25 degrees. It is difficult to say that B is at fault here if AI's action is not the result of some bad discrimination, especially in the case of a more advanced system where the AI monitors the vital parameters of the spouses and "knows" that, for instance, spouse B's blood pressure may be very low that day.

The second type of system will be an autonomous servant of only one user. This may result directly from the architecture of the program, from the content of the

license agreement, or from the nature of the software installation and set-up. If, therefore, one of the household members exercises power over such an autonomous system, the will and preferences of the others will either not be taken into account at all, or will be taken into account to a narrower extent; in the latter case, the extent will be determined by the will of the household member "in charge" or by the legal regulations and specific factory settings of the system.[19] The legal qualification of events resulting from the operation of such AI is not at all obvious. Seemingly, it could be considered that the system transfers or rather reflects the will of the user; as such, he should bear responsibility for the effects of such actions, which will consequently be evaluated in the same way as in the "analogue" world. In reality, however, the AI, does not so much carry over or reflect the user's will, but rather pursues the user's interests as it 'understands' them, albeit within the limits set by the design of the system; although it is important to note that in such circumstances, this AI would be an automatic system rather than an autonomous one. Regardless, an ideal system will mimic and anticipate the user's will and act in accordance with such a hypothetical will—this action will not always be in accordance with what the user himself would actually do. On the basis of currently binding civil law, it will not always be possible to attribute the effects of such an action to the user. The mere fact that an entity obtains a benefit from a specific action by another entity (more broadly: any action, even accidental) cannot be the basis for a charge necessary to ascribe liability for damages on the basis of fault. In turn, unjust enrichment (unjust enrichment, *ungerechtfertigte bereicherung*) and strict liability may come into play if specific rules exist in that regard (e.g. in connection with the movement of an enterprise). In the domestic context, it is difficult to identify *de lege lata* such norms: the householder is not liable on the basis of strict liability for consequences resulting from accidental events originating from his objects, unless some special situations are involved, e.g. related to the ejecting, effusion or falling of an object from the premises.[20] If the AI system is considered as an object that causes damage, any liability by the owner (user) may be based on the principle of fault. However, any link between an autonomous decision by an AI to fault seems a tenuous one. It is also difficult to consider that the mere use of such a system is culpable; culpability could in such a case be attributed only in cases where the user used the AI in an improper manner or did not make appropriate checks, for example.

In the case in question, AI is acting for (or "on behalf of") user A, but it is also acting for user B; however, the specific action is more in line with B's interests than A's. As such, the real problem is that A has no claim on the AI. One might say that the AI has ignored his hypothetical will or interests. However, it is difficult to hold

[19] For example, it is very likely that such factory settings will incorporate safeguards related to the protection of human life and health. In such a case, e.g., a system that realizes only the interests and will of person A may react in the interest of person B to a certain extent, e.g., upon noticing such person's fainting, contact the ambulance and open a window.

[20] The liability for the damage inflicted by the ejection, effusion or falling of any object from the premises which is modelled upon the Roman action *Actio de effusis vel deiectis,* functions in some countries, for instance § 1318 ABGB (Austrian Civil Code) and Article 433 of Polish Civil Code.

user B responsible for such an action. As such, the situation is only seemingly identical to that in the analogue world: if spouse A behaves against the will of B and changes the status quo, he can be liable to B. However, if the autonomous system performs the action, A does not bear responsibility. In such cases, a claim may be brought on the basis of unjust enrichment laws, which are detached from fault and indifferent to the source of enrichment (which could also be the action of an AI). However, such a possibility will come into play only where there was enrichment at the expense of another person, and no such situation occurs in the analysed case.

Finally, in the case of the third type, the AI could tune in simultaneously to all household members and pursue their common interests. However, such an AI must itself resolve conflicts of interest that naturally arise between users. This would most likely involve relentlessly searching for balance points between these interests and tuning to subsequent data, based on Bayesian inference. Here too, however, conflicts will inevitably arise, as mentioned above. This system is potentially the most dangerous and the most difficult to frame, as it is possible that the AI may take over a specific sphere of activity of household members. Indeed, such a state of affairs would be more likely in the case of more advanced systems, with a broader scope of control. In such cases, it will be the AI that decides on the conflicts of interests between all users, and as they will all share equal positions, there will not be any easy mechanism to assess and verify the will it ultimately expresses. In fact, the elimination of a decision made by the AI would require the consent of all users: a change by one of them, even if possible (legally and in fact), would mean entering into a potential conflict with the rest of the household members, which is, after all, part and parcel of life in any family. However, more importantly, it would also mean entering into a kind of dispute with the AI, which, as can be expected, will make decisions with a high level of accuracy when attempting to balance the users' interests. It is therefore possible that any desire to challenge the decisions of the AI, requiring active action that may be potentially questionable in the face of AI's efficiency, will be more of an exception: for purely pragmatic reasons, and for the sake of quiet life, the household consensus will likely be to take the AI's decision at face value and not constantly verify it. This is, after all, what such systems are meant to do: make life and work easier and more convenient. This, in turn, will simply lead to the family relinquishing power to the electronic hands of the machine. Let us note that such effects will occur in each of the three types of home AI systems given above; however, in cases where the AI is subject to the authority of a single person and where it realizes the will and interests of this person, any loss of control over one's own home life (or its specific aspects) will be more difficult, or rather more subtle; after all, the operation of the system as a whole will be specifically subject to the will of a single household member. In the case of the third type of system, it becomes autonomous on a meta-level: it is no longer only independent of the direct commands of the specified user, but it is independent of all users, since it must itself examine, evaluate, balance and implement the various preferences, interests and hypothetical wills of the household members involved. This averaged will is not the will of anyone but the AI itself.

When assessing the effects of certain decisions made by the AI on the legal-family or legal-estate relations, or even on the obligations of household members affected by the system, one cannot help but notice the complete inadequacy of existing normative solutions. Such cases will doubtlessly entail unexpected problems, with no immediately obvious solution for the law, or more precisely, the court. The paradox outlined here will consist in the fact that *de facto* (because of course not *de jure*) it is people (here: household members, but the concept can be successfully transferred to various social relations) who will become executors or at least passive acceptors of the machine's will, while it should be the other way round. This may even lead to an actual reversal of roles: the AI will become the decision maker, while humans will be subject to its will. It must be emphasized that such a state of affairs will probably not arise in the legal sphere, but the example of multi-entity family relations shows that such subordination will be inevitable; it will eventually come true, although no one will expect it. Going against the will of the AI, which will most likely make more rational and balanced decisions, will inevitably raise the question of liability for damages for actions contrary to that decision. Of course, this is not a question of liability to the AI, but to other entities that will be able to, probably convincingly, demonstrate that any behaviour contrary to the AI's "suggestion" was suboptimal, at least on a level where it can be monetized. As such, the course of action suggested by the AI will eventually become the only choice available for its human users, because its rejection will be accompanied by at least an indirect sanction of compensation.

Another case would be where one of the users is able to impose his or her will on the others. This could take place in a wide range of circumstances depending on the system configuration and the specific factual situation. It is possible, for example, to imagine a system which, when first set up, is configured with only one user and, for example, responds only to his verbal commands. It may also be that the system tunes itself to all users, but it has been used for so long by only one person that it only implements his preferences. Again, general principles govern how certain facts are viewed in the analogue world. For example, take the case of a system that chooses music or films based on user preferences. If the choice the AI makes reflects the preferences (interests) of one of the household members, then assigning blame (and therefore responsibility) towards the other will only be possible if, preferences that suit only one of the household members had been included in the AI configuration or training stage.

On top of the relationships between household members in the face of AI, another layer of legal assessment concerns the issue of the collection and processing of personal data which, of course, is ruled by GDPR, but also bears problems going much further. The use of AI-equipped objects in the household will involve the collection of personal data. In fact, it is likely that such systems will perform better as they collect increasing amounts of diverse data. Even so, this potential must remain in line with the requirements of the law. A user who sets up an autonomous system at his home that collects information about him must give his explicit consent to do so. However, the matter becomes more complicated when you consider that such a system may also collect the data of other household members. Such collection

should require their consent, but at least in some cases, it will not be easy or even possible to separate the data of certain household members, particularly data based on householder behaviour, which is only obtained by indirect tracking. In addition, one of the users may deliberately collect, and use, personal data of other household members, or use data collection tools to track them. Again, the principles remain the same: the possibilities for collecting personal data are greatly facilitated and expanded through Internet of Things systems, but the legal environment in this regard remains the same. The mere fact that the user is using a smart system does not allow the user to do more than with more traditional means; on the contrary, because the system can collect more personal data, in a faster and more varied way, the user in fact has a greater, rather than a lesser, obligation to ensure that such action complies with the law and the rights of those individuals, led by the right to privacy.

The difficulties described above as related to the household may also appear in other situations when AI actions embrace a social group connected with not only legal but also personal bonds. On a larger scale, such groups can be parishes, schools or cities, for example.

8.5.2 Joint AI in a Company

Particularly complex problems may arise when an AI is part of an enterprise. Although such situations are already relatively common today, the problems that arise from it are yet to be identified. In our view, the same phenomena associated with the pursuit of diverse interests in a household may also arise in the enterprise; however, their detection and understanding may be more complicated.

The will of a legal person, especially one with a complex structure, is not constituted merely by the simple sum of the actions of the persons comprising its body. Decisions made collectively ("collective will") are somehow the resultant of the interests at stake. The use of AI, at any level of business management, can result in certain decisions being taken entirely (or to a large extent) out of the hands of humans. In such cases, it is the AI that will assess how to balance the various interests and reach a point of balance, and this may be different from that which would be reached by humans. The use of such a system may, moreover, involve the assumption that decisions will not only be faster or cheaper, but also better, i.e. that they take into account more data and more effectively pursue the identified objectives. There are growing demands to make company law more flexible with regard to how the function of the company's body is being taken over by AI (or possibly one of the members of the company's body). This may take place through changes in the provisions concerning existing companies (more broadly—legal persons), but it may also be carried out by creating a new type of legal person, in which the function of organs (all or some of them) will be performed by the AI.

The point of both introducing AI as an assistant to those who sit on the bodies of legal persons and as a member of such bodies is the same: it is about improving the

operation of that body and the organization as a whole.[21] However, "improvement" itself may be defined in different ways, as the objectives and principles guiding individual legal persons may also differ. Moreover, general goals, such as achieving shareholder profits, may be different from more immediate direct and specific goals, such as finding the best contractor for a particular investment. If the immediate objective is to balance the interests of specific actors (e.g. shareholders or employees), the balance point reached by the AI will, usually, be better than that potentially found by people. This can have at least three dangerous consequences:

Intentionally or not, AI may be trained to favour certain interests. This may be so covert that it is invisible to the company as a whole. This preference may result from various reasons, the most obvious being that the use of particular data during the AI development stage. The problem of discrimination in decisions made by AI is already widely discussed and it is beyond the scope of our discussion to analyse this further; suffice it to say that the system should be designed so that discrimination (and hence preference) does not occur. However, let us note that it is also not excluded that AI will pursue some interests of its own, separate from the interests of anyone associated with the legal entity itself (board members, shareholders).

However, it is not excluded that, regardless of the choice of data, the AI will independently find a point of equilibrium between the interests at stake, which may appear like a preference (or discrimination) in relation to an earlier (human) action. After all, it is inevitable that the point of equilibrium will be different from the previous one, and thus some interests will be satisfied to a lesser extent than before. Consequently, it will not be AI that will prefer or discriminate, but rather that through its action, it will reveal prior discriminations or preferences.

The assumption that the choice made by the AI is optimal implies that human decisions inconsistent with the AI's decision or that were made without the AI are suboptimal, and raises the question of whether they should be permitted. In such cases, the acceptability of such an action will have to be determined at the level of the internal acts of the legal entity itself. However, it can be assumed (as mentioned in Chap. 11 when analysing negligence) that the will of the AI will usually be binding on humans, if only for the reason that deviating from this decision will be difficult to justify based on rationality or effectiveness when achieving the set goals. This may entail liability for negligent action (e.g. of the board of directors). For example, shareholders interested in the best possible development of the company may not be interested in the details of the decision-making principles, but they may feel dissatisfied by the information that an inferior solution to that proposed by the AI was chosen. While this may give rise to some kind of liability for damages in the legal

[21] Such a practice is not a science-fiction. For several years there have been different instances. Cf. S. Sharwood, Software 'appointed to board' of venture capital firm, "The Register" Sun 18 May 2014, https://www.theregister.com/2014/05/18/software_appointed_to_board_of_venture_capital_firm/ J. Bates, I'm the Chairman of the Board, "Huffpost', 06. 04.2014, https://www.huffpost.com/entry/im-the-chairman-of-the-bo_b_5440591, last access on the 2022. Cf. also Armour and Eidenmueller (2019).

sense, it may also result in the exclusion of the human factor in cases where it is be ineffective.

The problems which appear when AI is acting within the company are on the one hand easier and on the other hand more difficult than those existing in the household. They are easier because there are no personal obligations, moral or legal, between the humans inside the company, but they are more difficult because the economic interests which are at stake may be of the much greater value. This may be a reason why in certain situations, the users of AI may demonstrate more compromise and co-operation, while in others, they may be less accessible. However, it seems that an AI system, as an unconscious and unemotional instrument, may be better suited to tackling the situations in which a solution may be calculated according to a numerically-assessed value.

8.6 AI as a Household and Family Member

Even more difficult problems than those outlined so far involve the direct interference of AI in personal relationships between household members. This interference into the very essence of human relationships, which will inevitably become increasingly frequent, must affect both these relationships and their legal assessment. This is closely related to the issue of personal interests, which are explored in Chap. 6.

One of the fields that represents a potential challenge concerns that of spouses and other persons close to them. These are bound by various ties, some of which are provided with a direct or indirect legal sanction, such as the obligation to care for one's spouse or to raise a child. At least some of these duties could be fulfilled to some extent by an appropriately constructed AI. An autonomous system, and sometimes a robot, could (technically speaking) satisfy some of the emotional and physical needs, and could provide care for a child or partner.[22] However, it is questionable whether this kind of "assistance" by a machine could be considered sufficient from the point of view of legal obligations. As certain systems develop, the boundaries between performing duties in person and performing them using

[22] So-called *social robots* are becoming increasingly popular, although they are still imperfect. For example iPal Companion Robot, a 3.5-foot-tall robot, is promoted with the description: "The iPal Companion Robot is a teacher for children with spoken language learning and tablet-based educational programs, providing educational content in an engaging manner that supports social development and encourages interest in science and technology. iPal also makes education fun and appealing for children. It can talk, dance, tell stories, play games, encourage physical activity, and enable them to chat with friends, share videos, and safely connect to the internet and social media. Parents, under strict controls, can remotely control iPal and monitor their child's progress, safety and activities on their smartphone or desktop from anywhere and at any time. Many elders are alone and lonely. They often have problems keeping track of everyday activities, such as taking their medicine. iPal is a constant companion that supplements personal care services and provides security with alerts for many medical emergencies such as falling down". https://www.robotshop.com/en/ipal-companion-robot-blue.html, last access on the 4th of August 2022.

technology will become blurred. The very "personal" nature of certain activities or obligations will be redefined, due to the permanent assistance of AI. The doubts this raises in contractual terms are discussed in Chap. 9.

In our opinion, it cannot be ruled out in advance that certain personal relationships, and consequently, such rights and obligations of this nature, may be assisted or even replaced by AI. A suitably-trained AI may create such a perfect avatar of its user that it can successfully replace him in certain situations such as sending messages, which can be even more complicated than "I love you". Of course, when the origin of such a message is unclear, its value drops dramatically. It is an open question as to what extent such behaviour can be considered legally relevant. After all, from the point of view of legal-family relations, as assessed by an outside observer, it may play an important role in how certain obligations are fulfilled, particularly in connection with the settlement of divorce or child custody cases. Additionally, at least some such AIs may fulfil certain emotional needs expressed by family members, and hence could not be viewed as a mere object of some property law rules. The realization of a social function, i.e. as a family member, will result in the anthropomorphisation of such AI to an extent that impinges on ownership rights. The owner (user) of the AI would be limited in its use, and even in the right to modify or disable it, if such activity were to affect the personal interests of family members, as well as their feelings and relationships.

The further development of the AI will therefore make it necessary to remodel the regulations relating to the family relationships in which this AI will be used.

In addition to the situations above, let us note one more of considerable importance: making decisions of significant importance for other family members, such as those relating to the custody and representation of the child. Can an autonomous fridge decide whether a particular product should be given to a child, and at what time and in what quantity? Furthermore, in terms of representation, the following example. A certain AI with medical specifications allows early diagnosis of certain diseases, predicting their development and planning correct treatment. What is the relevance of its "advice" to the child's parents? Can the AI itself consent to the child's treatment (including its surgery)? The answer to this question requires us to assume that, bearing in mind the large number of parameters relevant to the child's interests, the AI's decision may be the most rational one; in fact, the odds are that this will almost always be the case.

This, by the way, is a reflection of a broader problem previously noted regarding the accepted standard and quality of decisions. If the given AI reaches a superhuman level of professionalism within its domain of action, e.g. paediatrics, its indications will also be worth more than evaluations of experts, e.g. physicians. The difference, of course, is that caretakers protect the whole range of interests of the child, not just those related to the performance of one specific procedure. As such, an AI may consider that procedure necessary, but the child's caretakers may consider it unjustified for other, non-medical reasons. If so, this kind of AI could only have an auxiliary function, perhaps as an advisor in a certain, narrow scope, but it could not replace the parents in making the final decision. Of course, the same applies to replacing an adult's decision with that of an AI. When making a decision to receive

treatment or not, each person decides not only about his or her health and protects his or her other interests going beyond the medical issue itself, but above all protects his or her autonomy and thus dignity. It is another facet of the more general problem on the intersection of human autonomy and the autonomy of the AI, which cannot be ignored because—at least to a certain extent—it will have a better capacity to make appropriate decisions.

8.7 AI as a Possessor of Property

Due to the fact that AI makes autonomous decisions, it should be considered whether it may be treated as a possessor of a thing. According to the civil law tradition, a possessor is a person who in fact wields power over a thing (*corpus*) which is connected with the will to exercise this power for oneself (*animus rem sibi habendi*). If the one who controls the thing does it for someone else, he is a dependent holder (*animus possidendi pro alieno*).[23]

Both the question of wielding power over a thing (*corpus*) and the manner of exercising this power (*animus*) in the case of AI appears to be a new and not obvious matter. What does it mean to be in possession of a thing? It seems that what is meant here is an actual state in which a given subject (or more cautiously: entity) has an actual possibility of making decisions with regard to the thing (irrespective of legal relations concerning the thing), i.e. "holds the thing in his hands". This conception of wielding power has been—at least in its core—accepted since Roman times. The ability to decide on a thing, to take certain actions towards it, is therefore crucial. Wielding power is not excluded then by the fact that the one who wields power is not independent in executing it, because to some extent he must take into account the decisions of another person. The holder is dependent on the dependent possessor, and the dependent possessor is dependent on the autonomous possessor.

As wielding power is a certain factual state with only specific legal consequences, the problem of power and possession has been considered only as one that concerns legal subjects. In other words, as they have no subjectivity, it is impossible for some things, such as animals, to control a thing in the legal sense, although it may be possible in a purely factual sense.

This would seem to be a perfect time to take a new look at the issue of legal subjectivity (as discussed in Chap. 2). In light of the findings of the chapter on subjectivity, it can be seen that the social context is one of the main premises upon which the law maker awards AI with subjectivity, at least to the extent that it 'wields' a thing. In other words, the fact of an AI deciding about a thing is so significant socially that it cannot be ignored by the law. For example, the fact that a car is driven by an AI cannot be ignored by the law, because it would lead to irremovable paradoxes, i.e. it would require the acknowledgement that the vehicle is driven by

[23] Dias (1956).

someone other than the AI: after all, the most obvious manifestation of wielding power over a car, at least during road traffic, is driving. If one assumes that AI does not possess the vehicle, then it should be counterfactually assumed that it does not drive it; on the other hand, if one assumes that the AI drives the vehicle, then it does have power over it.

Therefore, can AI be considered as wielding power over a thing in the context of the conceptual grid of property law? We believe this to be the case: such a qualification is the most adequate for describing the situation in which an AI controls a thing. Since it acts autonomously, and its control of the thing is not merely a transfer of the will of someone else, it has power over the thing, i.e. it holds a thing to some extent. In an autonomous vehicle, the AI performs the function of a driver, who could be considered a holder, *vis-à-vis* the possessor, of the car. The same will be true for Internet of Things devices controlled by AI: in such cases, the AI wields the thing, i.e. it makes decisions relating to the thing in a peremptory manner, without anyone's direct control or approval.

But is it reasonable to speak of *animus* in the case of AI? And if so, what is the nature of this power ?

There is no doubt that weak AI does not have *animus* as understood in the case of humans. AI software, or even embedded software in a robot, has no intellectual or emotional relationship to its position in the world, nor with the thing it controls. However, civil law has typically paid little attention to the mental attitude of the person wielding power over a thing towards it; it has focused more on the way such person behaves, i.e. how his or her manner of ruling can be socially positioned. In order to determine the *animus* it is not important what the driver of the car thinks, but how he behaves: what counts is the entire network of legal and factual relations in which he remains. It is on the basis of these objective assessments that we can determine whether he behaves like an owner (and is an autonomous possessor), as a tenant (and is a dependent possessor) or perhaps only as a person who holds the property for someone else (and is a holder). There is therefore no obstacle to attributing *animus* to legal persons and, it seems, no obstacle to doing so in relation to AI.

An AI that drives an autonomous vehicle or any other thing can, in principle, be qualified as a holder, i.e. one who wields the thing for someone else. As the AI has no interests of its own, nor does it pursue them, it is not possible to qualify this power as possession.

In certain situations, however, it is also not excluded that AI can wield some goods for itself; such a qualification would, of course, be possible only if it were legally permissible for AI to be entitled to certain goods, i.e. to be an owner. If it is assumed that AI may acquire a certain degree of subjectivity, e.g. in relation to monetary amounts used to make transactions, as noted below, it is obvious that AI can be qualified as an autonomous possessor of these funds.

The classification of the power of an AI over a thing as possession (possibly as holding) is of fundamental importance for property law and more broadly for the whole of civil law. Most importantly, it is connected with the prohibition of infringement of possession: the prohibition of lawlessness (arbitrariness). However,

this prohibition must be understood in a different way for AI than in relation to humans. It is at this point that the context of property law collides with the problem of human autonomy. While it is clear that the power of AI over a thing must be protected, otherwise such a construction would not make any sense, the scope of acceptable protection against human arbitrariness available to an AI must be particularly limited in connection with Asimov's first law, already cited.

8.8 AI as an Owner of AI

Since an AI could be both a possessor and an owner, the question arises whether an AI's property can include another AI (of course, such a storied construction could be extended). In turn, the answer arises as a consequence of the assumptions made earlier: there is neither a general capacity for AI to own another AI, nor is there any obstacle to a particular AI having such a capacity. It depends on the purpose of its action and the subjectivity and legal capacity granted to it to achieve that purpose. In principle it can be said that such storied constructs can be very difficult to supervise, and explaining the inter-relationships can be too complicated for humans. This therefore risks the potential loss of real control over the operation of AI. It seems therefore that the possible admission of such dominant and dependent AIs should be strictly limited.

Let us look at this with a hypothetical example. A sophisticated system operating in some field of commerce uses its own resources to achieve a well-defined goal. As part of this purpose, the AI is allowed to make use of other AIs, available on the market, which can improve certain aspects of the action being undertaken, and which are not feasible for the original AI itself. Within the available resources, the AI therefore enters into appropriate licensing agreements allowing it to use these "subcontractors". The user (operator) may not even know that he is "using" their services—in fact, he does not need this knowledge. In an economic sense, of course, it is he who will benefit; however, in a legal sense, the AI of the second level will belong to the AI of the first level.

8.9 Succession of Rights of AI

The assumption that an AI may be, to some extent, an owner (possessor, holder) raises the need to clarify the rules of succession when (1) the AI expires; (2) the AI multiplies itself.

The problem of succession must of course be linked to the clarification of when AI ceases to exist, i.e. when a possible succession could come into play. One can look at this issue from several angles:

A. The complete destruction of the code;
B. Destruction (change) of the code in such a way that a new AI, different from the existing one, is created;
C. Disabling the operation of the AI without destructing the code;
D. Deletion from the register or the arrival of another moment indicated in the register (i.e. legal expiry).

Re. A. The complete destruction of AI code seems unlikely today, but of course such a possibility cannot be excluded. In such a case, the AI will undoubtedly cease to exist, which inherently deprives it of legal capacity, if it had. Of course, the natural consequence of the destruction of the code should be the deletion of such AI from the register.

However, it may be more difficult to assess a situation in which there is not an annihilation of the parent AI, but of its specific, individualised version. For example, over time, the AI in a particular device becomes trained by the user to accommodate his behaviour and adapt its decisions to its credentials; as such, two initially identical AI copies will become sufficiently different that they would make different decisions in the same situation. The annihilation of one of the copies therefore does not affect the bases of the AI (the original code); however, it destroys the "superstructure" created in the process of interaction with the end user. After several years of use by different people, two autonomous fridges will order different wine for Saturday night. It is therefore important to consider the consequences of the destruction of this "superstructure". Note that the register describes the output version of the AI. "Superstructure" is no longer part of the system but has a close relationship with the user.

Re. B. In the case of continuously self-learning AI, the output code is transformed. This can also include the self-repair or self-modification of the code by incorporating elements of other programs. In a broad sense, this is analogous to human maturation or evolution. Determining whether it is still the same AI or whether a new AI is emerging will be both technically and legally difficult. It is not just a question of facts, but also of law, which must identify the properties that differentiate a concrete AI from already existing systems. The creation of such a new AI, also through evolution, may also necessitate an appropriate registration procedure.

Re. C. The operation of AI requires data processing. "Dead", i.e. not running code, is of course of no relevance. However, it can be a significant difficulty to define how AI in dormancy differs from AI that has been switched off entirely: whether switching off is meant to imply a permanent cessation of operation and, at the same time, to indicate what permanence means in this context.

Re. D. The above-mentioned difficulties lead to the conclusion that the possible succession of AI law must be preceded by a legal "declaration of death", which—as it seems—cannot be expressed in any other way than from the normative point of view, i.e. legal definitions determining when a given AI is to be regarded as deactivated (or destroyed, annihilated or put to sleep) must be created, and these must be associated with the relevant registration obligations. In other words, we are

of the opinion that the law should recognize that the 'death' of an AI is a legal phenomenon that occurs in a certain way. Let us note that we have already prejudged that only a registered AI can act in the market. Therefore, striking it off the register deprives it of its legal capacity. In this view, it is not important whether it can exist outside the registration: even if technically possible, such existence would be extra-legal.

It is only when the law determines that AI as an entity ceases to exist that the question arises as to the possible allocation of the rights to which that AI was entitled. In terms of personal interests, as noted in Chap. 6, these rights should continue to exist and be subject to protection: the normative construction to anchor this protection may be potentially different. In relation to property rights, on the other hand, there is a need for the creation of a succession mechanism. Just as the "death" of the AI itself must be predetermined, described by legal rules and specified in detail in the registration, so should the possible allocation of its property. If the AI's property is to serve certain purposes defined by the operator (e.g. an autonomous fridge having its own budget), the eventual expiry must result in the acquisition of this property by whoever is the operator, or whoever has another right to the AI; these issues must be resolved when accepting the admissibility of AI having its own property.

The issue of possible multiplication (copying) of AI must be approached in a similar way. Each copy must be considered a separate entity, which does not become a legal successor of its original copy in any respect: it does not become a co-holder of the property that belonged to the original version. AI can therefore not be treated as a species, but always as individual examples: each copy is a separate entity whose endeavour may be different. This consideration is particularly relevant in cases where the AI continues to improve itself on the basis of successive experiences after being activated. Thus, the possible property rights of the AI are not transferred or divided in the event of code multiplication but remains assigned to the parent AI. Possible intellectual property should be assessed in the same way; however, this was discussed in more detail in Chap. 7.

8.10 The User's Death and the Succession of the Personalized AI

Together with the issue of the destruction of a personalized copy of an AI, there is a need to clarify the situation of ownership and trading regarding such versions of AIs. There is no doubt that it is necessary to comply with the rules concerning the transferability of certain software (including AI) to others, and that the principles of GDPR concerning the protection of personal data must be observed. However, it must be considered whether a personalized version of an AI, being a kind of shadow of the user (silhouette), constitutes a special kind of legal good, which is partly of a personal nature. It is important to note that the point here is not whether the AI will

store any personal data about the user, but that it may restructure itself in such a way that it becomes adapted to the user in the course of its interaction. The decisions it makes will be profiled and personalized, and will fit only to that particular user. Trading in such AI should, in our view, be subject to separate legal rules.

Note that (non-autonomous) machines carry and replicate the user's will, which is written (programmed) into them. Specific problems arise in the case of autonomous AI, i.e. those which can replicate the user and learn to carry out the user's hypothetical will, with respect to continued operation after the user's death. The hypothetical will of the user in this case is not what the user actually wants to do, but what he would probably want to do or would want to do if he thought of it at all. A person may be dead, but machines can act on his behalf as if he were alive.

The issue of possible declarations of intent are dealt with in more detail in the consent chapter: the problem can be solved by referring to analogous solutions relating to the effects of a possible further action of the proxy after the death of the principal. However, personalized AI can also act in a much broader dimension, realizing the hypothetical will of its user in the purely factual sphere. In the home, the system will adjust the temperature accordingly, and on the Internet, the machine may comment (as the user) on the latest political or family news in a characteristic way; such actions can influence other people to a certain extent, and can even cause conflicts of interest or other damage. Their effects should be attributed to the legal successors (e.g. heirs) of the user, as long as they have been given the power to disable or alter the action of the AI. The scope of such attribution may be the same as that appropriate for actions performed by the AI while the user was still alive. If we assume, therefore, that the AI developed its commenting style from the user's comments, should it continue its Internet interactions in a similar vein while infringing the personal interests of others, and that it acts "on account" of the user, who is responsible for it, it should be consequently assumed that even after the user's death, such an AI acts on account of his or her legal successors.

Another issue concerns the question of whether the existence of personalized AI may be acknowledged as an element of the personal interests of its user. Since such an AI-silhouette is an imitation of the user's personality, it cannot be regarded as purely a commercial object. This question could raise difficulties regarding the intersection of property rights between different persons or may be limited by the content or time and the immaterial rights of the user. This issue is examined in the Chap. 6.

References

Books and Articles

Armour J, Eidenmueller H (2019) Self-Driving Corporations? ECGI Working Paper Series in Law. http://ssrn.com/abstract_id=3442447, last access on the 4th of August 2022

Brown RD (2020) Property ownership and the legal personhood of artificial intelligence. Inf Commun Technol Law. https://doi.org/10.1080/13600834.2020.1861714
Cheng Y, Zhang X, Tang X, Zhu H (2021) Is AI better than human in identifying high-potential talents: a quasi-field experiment. AMCIS 2021 Proceedings. 13. https://aisel.aisnet.org/amcis2021/art_intel_sem_tech_intelligent_systems/art_intel_sem_tech_intelligent_systems/13, last access on the 4th of August 2022
Corrales M, Fenwick M, Forgó N (eds) (2018) Robotics, AI and the future of law. Springer, Singapore
Dias RWM (1956) A reconsideration of Possessio. Cambridge Law J 14(2):235–247. http://www.jstor.org/stable/4504402, last access on the 4th of August 2022
Dyrda A (2015) Why legal conventionalism fails?. Archiwum Filozofii Prawa I Filozofii Społecznej 1/10. https://doi.org/10.36280/AFPiFS.2015.1.14
Kaplan J (2016) Artificial intelligence – what everyone needs to know. Oxford University Press, Oxford
Katz A (2010) Intelligent agents and internet commerce in ancient Rome, society for computers and law. https://www.scl.org/articles/1095-intelligent-agents-and-internet-commerce-in-ancientrome, last access on the 4th of August 2022
Lanovaz M, Hranchuk K (2021) Machine learning to analyze single-case graphs: a comparison to visual inspection. J Appl Behav Anal. https://doi.org/10.1002/jaba.863
Lewis D (2002) Convention: a philosophical study. Blackwell, Oxford
Mirhoseini A, Goldie A, Yazgan M et al (2021) A graph placement methodology for fast chip design. Nature 594:207–212. https://doi.org/10.1038/s41586-021-03544-w
Pagallo U (2011) Killers, fridges, and slaves: a legal journey in robotics. AI Soc 26(4):347–354
Pagallo U (2018) Vital, Sophia, and Co. The quest for the legal personhood of robots. Information (Switzerland) 9(9):230. https://doi.org/10.3390/info9090230
Ravuri S, Lenc K, Willson M et al (2021) Skilful precipitation nowcasting using deep generative models of radar. Nature 597:672–677. https://doi.org/10.1038/s41586-021-03854-z
Unger RM (1986) The critical legal studies movement. Harvard University Press, Cambridge
Wojtczak S (2011) O niewspółmierności wartości i jej konsekwencjach dla stosowania prawa. Wydawnictwo Uniwersytetu Łódzkiego, Łódź
Wojtczak S (2013) The broadening of legal notions as a tool in the neutralization of values in law. In: Pałecki K (ed) (2013) Neutralization of values in law. Warszawa, Wolters Kluwer

Documents

Directorate-General for Internal Policies (2016) European civil law rules in robotics. Study for the JURI Committee. https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf, last access on the 4th of August 2022
European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088 (INI)), P8_TA (2019) 0081. https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html, last access on the 4th of August 2022
The Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). Ethics of Connected and Automated Vehicles: recommendations on the road safety, privacy, fairness, explainability and responsibility", 17 September 2020. Publication Office of the European Union: Luxembourg, https://op.europa.eu/en/publication-detail/-/publication/89624e2c-f98c-11ea-b44f-01aa75ed71a1/language-en/format-PDF/source-search, last access on the 4th of August 2022

Chapter 9
Contract

9.1 Introduction

When an AI is an autonomous entity to some extent, any contract concluded by an AI, with the use of an AI or performed with an AI, raises a number of questions for civil law about the adequacy of traditionally-developed constructions and concepts. There can be no doubt that, in principle, such a contract should be regarded as the same as those concluded among humans, without the use of AI. The content of the contract, the rights and obligations of the parties arising therefrom, the rules relating to the place, time or way of performance of the obligation, are independent of how the contract was concluded. It is irrelevant whether the parties negotiated the contract personally or acted through proxies, or whether they used electronic agents. Of course, the circumstances connected with the very manner in which the contract was concluded are not entirely indifferent to the relationship between the parties. Some may be significant, for example, the party's right to withdraw from the contract (as in distance contracts with consumers), but they do not affect the essence of the contractual obligation, in particular the subject matter of the obligation, i.e. the performance due. However, the introduction of an AI into the contractual relationship, when it is an autonomous entity external to the parties, alters this relationship to such an extent that certain provisions relating to the rules of performance of contracts must be changed or adapted to the new situation; it may even be necessary to consider which of the known rules remain relevant.

The concept of contract belongs to the canon of civil science and is intuitively understandable. The DCFR defines

> II.- 1:101: A contract is an agreement which is intended to give rise to a binding legal relationship or to have some other legal effect. It is a bilateral or multilateral juridical act.

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_9

9.2 Adequacy of Basic Principles

9.2.1 Freedom of Contract

Freedom of contract is one of the pillars of civil law. Hence, it is usually mentioned at the beginning of any acts or parts of acts which regulate contract law, for example:

> UNIDROIT: Article 1.1. The parties are free to enter into a contract and to determine its content.
>
> DCFR: II. – 1:102(1). Parties are free to make a contract or other juridical act and to determine its contents, subject to any applicable mandatory rules.
>
> French Civil Code 2016: Article 1102. – Everyone is free to contract or not to contract to choose the person with whom to contract, and to determine the content and the form of the contract within the limits imposed by legislation.[1]

As explained in the Consent Chapter, AI should generally act as an agent of a party. In this case, the freedom of contract would apply to the very party of the contract, and the fact that AI would act on behalf of this party, in the name or in the interest of the party, would not make the situation of this party better or worse; the situation would generally be the same is if the AI were not present.

This does not mean, however, that the introduction of AI will not, in the long term, cause significant modifications to the understanding of freedom of contract, at least in some markets. It may be, for example, that entering into a certain type of contract requires the use of an appropriate AI. Given that the autonomous operation of AI implies a shift of the point of decision-making from the user to the machine to various degrees, the freedom of contract of such a person can no longer be understood in a traditional way. A party may in such cases express a general (directional) willingness to enter into a particular contract, but the parameters of the contract will be determined by the AI, with no (*de facto*) other possibility of entering into this kind of contractual relationship; these details can often be key ones, such as the nature of the other party, the price and the subject matter itself. In other words, it is likely that such areas of trade will develop, which will be 100% based on the operation of AI, and the entry into a specific contract in this market will also be a decision of AI. In this case, it is no longer possible to talk about the traditionally-understood freedom of contract, as the parties will in fact no longer have any influence on the establishment of the contractual relationship and its content. Of course, and this should be strongly emphasized, only certain specific types of market will be involved, such as those concerning certain fields of stock exchange trading or the virtual market in mobile applications or computer games.

Example: suppose that a computer game X allows the use of an AI that, among other things, interacts with other AI players, the aim being to achieve the best result in a game through this cooperation; furthermore, an element of this cooperation is

[1] The English translation provided by J. Cartwright & Bénédicte Fauvarque-Cosson, http://translex.uni-koeln.de/601101/_/french-civil-code-2016/, last access on the 2022.

the conclusion of contracts for the exchange or sale of elements needed in the game. All this turnover takes place in fractions of seconds, outside the consciousness of the player, who simply uses the facilities provided to him *a vista* by the AI serving him in this way. The contracts that such an AI enters into, including A2A (AI to AI) contracts, escape the classic model of "freedom of contract", since *ex definitione* the player has no influence either on whether he enters into a contract, or its content: his will is limited to installing the AI and defining its parameters, which, due to the specifics of the game, may be very general.

However, there would be much more serious consequences for the principle of freedom of contract, if the AI were allowed to act as a person or quasi-person in its own name, but on behalf of a third party. This would be particularly serious if the AI could do so as an ePerson acting in its own name and on its own behalf, like a legal person. We are of the opinion that in such a case there can be no question that the AI can be given freedom of contract, but rather that its strict regulation should exist for its right to participate in the market. AI cannot be given a general power to enter into contractual relations: the choice of whether to enter into a contract, with whom and what content. The scope of activity of an AI system must be precisely defined in advance, and thus its possible sphere of action in trading must be controlled and limited. This is not about general restrictions, referring to an ePerson or AI in general, but about specific restrictions imposed on individual AIs. Whether or not an AI will be able to enter into contracts, what these contracts will be and what, if any, could be their permissible content, should be controlled by the state and reflected in the registration parameters. This corresponds to the issue of legal capacities and UCDs discussed in Chaps. 3 and 4. Acting outside these limits, seen as acting without legal capacity, would always have to be qualified as leading to the nullity of the contract.

9.2.2 Freedom of Form

The principle of freedom of form exists in most legal orders. The conclusion of a contract is not subject to any requirement as to form. For example:

> UNIDROIT: Article 1.2 Nothing in these Principles requires a contract, statement or any other act to be made in or evidenced by a particular form. It may be proved by any means, including witnesses.
>
> DCFR: II. – 1:106 (1) A contract or other juridical act need not be concluded, made or evidenced in writing nor is it subject to any other requirement as to form.

The question is whether such freedom is adequate in relation to AI. In our opinion, this is the case, but only up to a certain level. From the point of view of subject matter, of course, nothing changes, i.e. a contract may be concluded in any way, even if it is concluded by an AI. However, the functioning of AI requires that the correctness of its operation be verified (the demand of transparency); this will require some coordination with the AI register and will also include the examination of the

legal capacity of this AI. Moreover, the need to ensure explainability, as a condition of operating in the market, in fact necessitates the adoption of certain technical solutions documenting the conclusion of a contract.

The fulfilment of all these indicated requirements does not allow any consideration of freedom of form, and the use of AI will, regardless of possible formal requirements concerning the subject matter of the contract, require a specific form of the contract. Such a contract, even if it takes the form of electronic or even voice communication (e.g. an autonomous system communicating via a voice interface), will no longer be the same as when the contract is concluded by a human being. In addition to information relating to the contract itself, additional information relating to the operation of the AI will have to be communicated at the same time. It will not therefore be permissible to choose a form of contract conclusion that prevents the transmission of these necessary data, in particular those related to the registration system.

To sum up, with regard to AI contracts, it is to be expected that the principle of freedom of form will be modified in such a way that this freedom will be limited to only those forms that ensure full synchronization with the administrative registration system. In practice, therefore, it will be the administrative rules relating to the operation of the future register that will directly determine how AI will operate in the marketplace.

All this indicates that some new forms of concluding contracts would be needed, but they would be dictated to some extent by the technical conditions of the time.

A good example here may be the phenomenon of the *smart contract*. According to Szabo (2006):

> A smart contract is computerized transaction protocol that executes the term of contracts. The general objectives of smart contract design are to satisfy common contractual conditions (such as payment terms, liens, confidentiality, and even enforcement), minimize exceptions both malicious and accidental, and minimize the need for trusted intermediaries. Related economic goals include lowering fraud loss, arbitration and enforcement costs, and other transaction costs.

Despite its name, a smart contract is not AI, nor does it need an AI. As such, is not smart at all, and furthermore, according to the dominant view, it is not a contract, in the legal sense of this word. However, some researchers, such as Durovic and Janssen (2019), p. 72, believe that smart contracts are capable of being formed as legally valid contracts. Of course, there are certain differences between smart contracts and "normal" contracts, such as self-enforceability and unchangeability, which may challenge some consumer rights, or the way that smart contract use the programming code instead of natural language. These differences, however, cannot be treated as harmful deviations from the traditional contract model. Smart contracts perform some useful and needed functions, the evidence of which is their growing popularity on the market. Therefore, the law must not ignore smart contracts and law makers should find a way of locating them within the structure of legal institutions.

For good reason some researchers differentiate between smart contracts, smart contract code and smart legal contracts.[2] They see that the smart contracts may be perceived from a different perspective. For example, the law may understand the concept of smart contracts by recognizing them as the form in which certain juridical acts may be concluded. This idea seems to be reasonable because as a matter-of-fact, smart contracts may have different content and imitate different juridical acts known by different civil law systems, depending on what is encoded in the software.

As to the type of form, it would be *ad eventum*, i.e. a form for the sake of achieving additional legal effects. These additional legal effects would be precisely the unchangeability and self-enforceability of the contract. Still, the use of this form may be limited in various ways; for example, it may entail a special informational obligation, similar to those determined in the Directive on electronic commerce.[3] This direction may be considered also in concluding contracts by some types of AI. Their use may be also acknowledged as a *form ad eventum*.

9.2.3 Pacta Sunt Servanda

One of the basic principles of contract law is that the parties are bound by its provisions—as a rule, a party may not withdraw from a concluded contract, unless there are circumstances provided for this by law or in the contract itself.

> UNIDROIT: Article 1.3. A contract validly entered into is binding upon the parties. It can only be modified or terminated in accordance with its terms or by agreement or as otherwise provided in these Principles.

> **DCFR: II. – 1:103: Binding effect**
> (1) A valid contract is binding on the parties.
> (2) A valid unilateral undertaking is binding on the person giving it if it is intended to be legally binding without acceptance.
> (3) This Article does not prevent modification or termination of any resulting right or obligation by agreement between the debtor and creditor or as provided by law.

There is no doubt that, as a principle, *pacta sunt servanda* should also apply to contracts concluded and performed by AI. However, as we have already indicated (Consent Chapter), the mere fact that AI is involved in the conclusion of a contract may be considered sufficient for the party who does not use AI to withdraw from the

[2] Woebbeking (2019), p. 108.

[3] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), L 178/1, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031, last access on the 4th of August 2022.

contract, of course for a short period of time immediately after its conclusion; this is analogous to the measures used to protect consumers contracting with entrepreneurs in distance contracts or off-premises contracts, in accordance with Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights. The stability of contracts must in this case give way to the overriding objective of protecting the autonomy, supremacy and dignity of humans in their dealings with computers.

However, it does not seem reasonable, at least at this stage of AI development, to recognise a general right of withdrawal from an AI-generated contract, i.e. this would be a general breach of the *pacta sunt servanda* principle. Protective mechanisms, as for B2C relations, must not go so far as to make it possible to undermine a contract at any stage of its duration, solely on the basis of the involvement of an AI system in its conclusion.

9.2.4 Subjective Circumstances on the Part of AI

As we establish in this chapter, a registered AI, having UCD[4] and thus also legal capacity within a certain scope, should have been able to act when concluding and performing contracts. On the field of contracts, civil law often refers to subjective circumstances, such as knowledge or awareness of a party, or an entity related to a party, about a certain fact, or an intention (intent) of a party, anticipation of a fact, understanding of a certain state of affairs, or good or bad faith of a party. This raises the question of how these subjective circumstances can be understood in a case where AI is involved in the contract.

> UNIDROIT: Article 1.8. A party cannot act inconsistently with an understanding it has caused the other party to have and upon which that other party reasonably has acted in reliance to its detriment.

> **DCFR: II. – 1:105: Imputed knowledge etc.**
> If a person who with a party's assent was involved in making a contract or other juridical act or in exercising a right or performing an obligation under it:
>
> (a) knew or foresaw a fact, or is treated as having knowledge or foresight of a fact; or
> (b) acted intentionally or with any other relevant state of mind this knowledge, foresight or state of mind is imputed to the party.

How do we interpret, in the context of AI, that it 'knows' about something, which here means that it *foresaw* something? There should be no doubt that in relation to the operation of AI, such knowledge is relatively easy to identify: it is simply a set of

[4] Cf. Sect. 3.2.1 in fine.

data that has been transmitted to the system. It does not matter whether this data was used, or how; the system "knows" about a given fact, even if certain data was deemed by the system itself to be useless and did not influence the final decision. It should also be considered that the system knows about such facts which are not contained in the data but can be derived from them through logical or mathematical operations; however, the extent of this knowledge is limited by the capacity of the system to perform such logical and mathematical operations, in accordance with its specifications. Thus, if the data contained the information that two apples and two pears were delivered, and that these are fruits, it should be considered that the system "knows" that four fruits were delivered, as long as the addition of natural numbers is provided for in the algorithm.

The example above opens the way for determining that there may be a duty to know, even though the system does not. Such a formula (phrased in different ways but meaning something like: "knows or ought to have known", "ought to have known, judging the matter reasonably", or "ought to have known if it had exercised due diligence") is commonly used in civil law. In relation to AI, this could refer to those situations where the system did not have certain data but should have, or did not process it in a certain way, despite the fact that it should have done so (e.g. did not add up the fruit in the example above), based on its specification.

In the same way, it is necessary to understand "predictions" about the occurrence of a particular fact in the case of an AI system. If a system has certain data and is constructed in a way that implies that certain conclusions should be drawn from those facts, then it can be put as "predicting" those facts. What is crucial in this context, however, is precisely the connection between the possibility of attributing predictive capacity to a system and the registration description, which defines, among other things, the goals and mode of action of AI. In relation to humans, in principle, the ability to predict certain facts is assumed to be the same; apart from special circumstances limiting mental abilities (age, illness), all homo sapiens, in principle, have a similar ability to predict a certain range of consequences based on a certain set of information. For example, if a person throws a copper ball from the tower, it will fly towards the ground and not upwards, and if he approaches the cash register with purchases in a shop, he will have to pay for them. There is therefore an assumption that all people share a set of basic abilities to predict the future on the basis of specific data. This assumption constitutes one of the pillars of law; this is particularly true for criminal law, but can also influence other areas, such as tort law. Human beings make decisions by anticipating the consequences of their actions, and in any case, for the purposes of the law, it is assumed that they anticipate, or could have anticipated, these consequences, regardless of what science actually says on the subject.

However, there is no such single common platform when it comes to AI; not only because of the diversity of autonomous systems, but also because of their obvious differences from humans. So without knowing what kind of AI we are dealing with, we cannot say anything about what it can know or predict based on the knowledge available to it. This information must be individualised. Only a registration entry could indicate what a given AI can predict. For example, a chess program is not

trained to psychologically analyse its opponent. It does not predict what the consequences will be if it wins against a person who is depressed, or who has bet his entire fortune that he will win against a machine. In contrast, an autonomous psychological support system can already combine certain facts of human psychology in a way that allows certain conclusions to be drawn, and in such cases, it could be established that the AI predicted (or could have predicted) a certain fact. This will obviously be more difficult when predicting human behaviour related to the human psyche, than predicting certain repetitive behaviours occurring in commercial transactions.

Hence, it can be said that it is not possible to appeal to "reason" or the accepted standard of a "supplier acting with due commercial care" (*die Sorgfalt eines ordentlichen Kaufmann*) or a "reasonable person" in the case of an AI. For there is not currently any minimum standard that can be expected of any AI (regardless of its scope of operation), nor will there be one for a long time to come, and therefore there is no model common to all AIs. This, in turn, leads to the conclusion that models and standards must be constructed with reference to certain types of AI, and not AI in general. Let us consider at this point what criterion should serve as the basis for building a set of AIs for which a specific model should be formulated.

The essence of AI is to pursue specific goals; autonomous AI does this by making choices. In this sense, an AI is defined by its purpose. Its scope of action is the description of the field in which these goals are pursued by the AI. A specific expected standard of action, i.e. a model that can be referred to when examining the 'behaviour' of AI, must therefore be constructed separately for each goal to be pursued, these goals being understood as the actions taken in a specific field. A particular level of diligence, and of expected "reasonableness" or reliability (if such phrases with subjective overtones are used in civil law), can only be referred to a certain category of AI pursuing a given purpose.

This problem may be clarified by the following comparison. When the law acts in the context of people with certain mental disabilities, it may require a certain level of reasonableness, foresight and diligence. However the mere fact that a particular AI provides some level worse, or not worse, than other AIs operating in that field gives no indication about whether that level is sufficient for operation in that field.

The purpose of AI must of course be described at the time of registration. The registration system would thus become the basis for the construction and understanding of subjective elements appearing in the hypotheses of legal norms. These norms date from before the time of AI, and refer to typically human properties. Of course, what has been said only concerns the understanding of the nature of subjective concepts in the context of AI; how the law would relate these AI properties, defined as such, to the level of expectation, i.e. the requirements, is another matter.

However, what forms a coherent thread in this case is the link to registration. For if a given field is characterised by an expected general level of a given subjective element, for instance knowledge or diligence (regardless of who would realise it, i.e. humans or AI), then an AI must undergo a check that this level is provided before it can be registered to operate precisely in that field. However, it is important to note that while an AI is not currently expected to perform at the same general level

available to humans (current AI is weak AI), it would be expected to demonstrate comparable performance, or even higher, for the subjective elements that are relevant to the realisation of the described goal. The model expected from AI will therefore always be a concrete model, never a general one. In summary, AI would be expected to be reasonable, to act according to the model of a "diligent merchant" or a "diligent doctor", but only to the extent needed to realise a specific goal; there is no general model of "diligent or reasonable AI".

Of course, before a given AI can be allowed to operate, these specific subjective features must be determined; however, the methods used will indirectly influence the overall determination of the expected level, and thus, the level required of humans acting with the same purpose. This issue is further discussed in Sects. 8.4 and 11.5. At this point it can only be noted that if the bar is set at a higher level for AIs than for humans operating in the same field, i.e. for the same purpose, there may be a natural displacement of humans from the activity in question due to the impossibility of achieving the same standard.

Sometimes, civil law uses the term *awareness*; however, in essence, this term can be regarded as synonymous with *knowledge*: if the creditor knew that. . . .= if the creditor was aware that. . . . These wording differences are not, in the intention of the authors of the legislation, semantic and meaningful, they are rather different forms of describing the same state of affairs. In the case of a human being, the condition for attributing knowledge to a certain fact is always the assumption that the human being is aware of her knowledge, and in general is in a mental and physical state that such awareness is possible. In the case of AI, of course, the issue must be described differently (*cf.* the section on legal capacity). Instead of awareness or consciousness, in this context, the UCD must be sufficient; for an AI system, being aware of a certain fact means therefore only that it is capable of doing so (has UCD) and has collected certain data.

As noted in Chap. 3, the appeal to (humanly-understood) consciousness in the hypotheses of legal norms is in fact a legal demand that the entity has the capability of attributing certain consequences to a given event. In relation to AI, the humanly-understood consciousness required by the law can be regarded as explainability, i.e. the possibility of explaining a certain action made by an AI; such knowledge allows us to ascertain whether a given AI has the capability to correctly attribute consequences to facts and make a justified decision on this basis.

In an analogous way, we can also examine the issue of the "understanding" of a circumstance by an AI. However, as mentioned above, such as cognitive process is not the same, or even comparable, to the human one, and will remain as such until AI becomes conscious.

> UNIDROIT: Article 1.8 (inconsistent behaviour) A party cannot act inconsistently with an understanding it has caused the other party to have and upon which that other party reasonably has acted in reliance to its detriment.

DCFR: II. – 9:102: Certain pre-contractual statements regarded as contract terms

(1) A statement made by one party before a contract is concluded is regarded as a term of the contract if the other party reasonably understood it as being made on the basis that it would form part of the contract terms if a contract were concluded. In assessing whether the other party was reasonable in understanding the statement in that way account may be taken of:
   (a) the apparent importance of the statement to the other party;
   (b) whether the party was making the statement in the course of business; and
   (c) the relative expertise of the parties.

Is AI able to understand anything? Can AI act in confidence in what the other party has stated to it? The issue requires careful consideration; the subjective elements and their nature must be understood differently for AI and for humans, and even between different AIs. Thus, it is only the knowledge of the type of AI, and the scope and manner of its operation, that will make it possible to ascertain *in concreto* what a particular AI can "understand", and identify the conclusions that can be drawn from this for the evaluation of a particular contract.

This raises the question of whether it should be acknowledged (presumed) that AI understands at least the same things that a human being would do in a given situation, i.e. in a given contractual relationship. Such a presumption, which is of a purely normative nature and does not relate to the actual design of AI, would be of significant importance for legal transactions and would allow us to escape from having to analyse the ability of a particular machine to "understand" the other party. Such escape would be justified by the fact that the admission to act in a certain field (e.g. contracting) must be linked to the fulfilment of certain requirements, which cannot be lower than for humans. In turn, a machine that does not meet such requirements, and thus, for example, would not be able to "understand" the statements or behaviours of the other party in a specific context, would not be able to act in a given field (its legal capacity would not include such activities). Consequently, a mere admission to act would imply that an AI has the capacity to "understand" to the extent necessary to act on a given field (or perform given actions). This presumption would include the ability to "understand" at a level no lower than a human, if humans are operating in that field. However, for example, if no humans were active in a given area and the AI was alone, the presumption would have different significance: it would entail a level of understanding established in that field, and this standard would not have to be in any way related to humans—it could be lower, higher or completely different.

The issue of AI-induced expectations in the other party gives rise to even greater difficulty. Although its interfaces may be designed to imitate human communication, an AI actually has no emotional or cognitive common denominator with humans: we share no common field that can allow us to assume that statement A will obviously elicit reaction B. In fact the content and form of an AI's statement is determined by its parameters. Furthermore, whether it can predict some further reaction or

expectation of the other party, and to what extent, also depends on the parameters of the machine itself.

In addition, when people come into contact with an AI, some may personify the AI and develop the same or similar expectations as they may have towards humans, while others will modify their expectations. The law must therefore insist on understanding these expectations completely objectively. Because statements made by an AI are analysed by a human according to his cognitive abilities, it is not the human who should adapt his thinking or expectations to the AI, but the machine should adapt to the human. Thus, a human, i.e. the recipient of an AI message, must not be disadvantaged by the fact that the message comes from an AI instead of from another human. So, if the AI's message creates certain expectations or trust they cannot be interpreted or assessed in the light of the parameters of the AI: while the parameters may be considered when the relationship is A2A (AI to AI), they must be assessed and measured by general human standards in the case of A2H (AI to human).

The issue of knowledge is naturally linked to the issue of good faith. Indeed, it remains to be clarified how to understand good faith in cases where it is supposed to be an attribute of the acting AI. The principle of good faith (*bonne Joi, Treu und Glauben, redelijkheid en billijkheid*) plays a central role in contract law. All civil codes adopt a provision on this principle. Analogous examples can also be found among the model rules:

> **DCFR: I. – 1:103 Good faith and fair dealing**
> The expression "good faith and fair dealing" refers to a standard of conduct characterised by honesty, openness and consideration for the interests of the other party to the transaction or relationship in question.

It is clear that the use of AI cannot in the least justify a lowering of the good faith standard. Such a general observation, however, does not preclude any assessment of whether the operation of AI does not cause that standard to be understood differently. It should be noted first of all that the definition of *good faith*, not only in the cited provisions, refers directly to the qualities attributed to people (e.g. honesty). This raises the following questions:

(a) how can these statements be understood when AI is at work?
(b) whose "good faith" is to be assessed in the case of an AI action—the AI itself or the user?

Re. a) With regard to the concept of good faith in the operation of AI, the primary focus should be on the registration elements of an AI. A certain ethical standard should be required as part of the registration procedure, and certainly its certification. The registration of AI thus provides a basis for formulating a presumption of good faith on the part of such AI, and this basis is in fact, stronger than in the case of humans or legal persons. The function of a particular AI, its objectives and mode of operation, as verified and described, should be consistent with good faith. Thus, if a particular AI acts, it should be presumed that its action is lawful, and this, *inter alia*,

presupposes good faith. However, such a presumption would be rebuttable (cf. Chap. 11) based either on demonstrating the falsity or incompleteness of the registration data, or on acting *in concreto* in a manner that goes beyond what we would define as acting in good faith in relation to human beings. This would include bad faith in a specific situation, e.g. resulting from the deliberate use of an intellectual advantage (cf. Chap. 10).

As mentioned above, if an AI system has some data, it "knows" about a given fact. If knowledge of a fact is a premise of some reasoning, i.e. it is an element of the hypothesis of some norm, then it should be considered to be true (satisfied) in that case. In particular, it may be sufficient to impute bad faith. For example, if an AI establishes that a purchased item was stolen, then the purchase of that item by that AI is made in bad faith; in a given legal system, this fact may be crucial for recognizing the invalidity of the acquisition. However, bad faith includes not only positive knowledge of a fact, but also situations where that knowledge should have been obtained. Should an AI therefore be required to act in such a way as to determine whether an item is stolen? The answer to such a question depends on a number of more specific factors, in particular the requirements imposed by a given law as to the exercise of care in purchasing a used item, and the context (circumstances) in which the contract was concluded. If, under a given national law, it is assumed that a person who knows, or by exercising due care could have known, about a given fact is acting in bad faith, then the issue shifts to the level of due care required of AI. Note that the issue here is not the diligence of the user, such as the owner of an autonomous refrigerator, but the diligence of the AI itself. The problem of due diligence and fault by AI is discussed further in the Chap. 11.

In the case of unregistered AIs, there is no presumption of good faith; on the contrary, bad faith should rather be presumed on account of illegal activity. Acting outside the registration system or exceeding the limits imposed by the registration should, as already mentioned, lead to the AI being deemed to have acted without legal capacity and, in those cases where this is of legal significance, to be deemed to have acted in bad faith. However, this does not exclude any possible proof of the contrary. This may be relevant in such legal systems where the notion of good faith is understood according to objective criteria.

Re. b) Attributing bad faith to AI means that its consequences will burden the user, even one acting in good faith. For example: the discovery that an autonomous system transacting in the art market has bought an object that it knows is illegally sourced would be sufficient to impute bad faith to the person who used this system.

The reverse situation, i.e. bad faith of the user against good faith of the AI, may be more difficult. Such a situation should be resolved in the same way as in the case of an agent acting in good faith for a person who is in bad faith. Such situations cannot be treated uniformly and their assessment may differ from one legal system to another. As a general rule, it should be assumed that bad faith on the part of the person using the AI in concluding a contract will lead to the contract being considered as having been concluded in bad faith, even if the decision was made by the AI. For example: if a user knows that image x is stolen and activates an AI to buy it, the conclusion of a contract to buy image x by the AI, even if the AI is acting in good faith, will not allow the acquisition to be considered in good faith.

9.3 Interpretation of Contracts Involving AI

It is not our task to address in detail the issue of methods of contract interpretation. However, we can say in general that in modern civil law (continental law), the most desirable and functional approach is the combined method, which includes both a subjective criterion, i.e. the common intention of the parties or the intention of the one-sided declarant, and the objective criterion, constructed by means of different-sounding general clauses. Any interpretation based on both criteria always takes place in a specific situational context, i.e. taking into account the circumstances surrounding the given contract. When interpreting a declaration of will, the actual will of the person making the declaration, i.e. making a unilateral legal act or participating in a multilateral legal act as a party, is taken into account, as well as his intention and purposes, and the trust and expectations that the declaration of will evokes in the other party or parties; with regard to these persons, the standard of a reasonable person placed in a given situational context is constructed.

The combined (or: subjective-objective, indirect, mixed) method has been adopted in most national legal orders. It has also been applied in both binding and non-binding acts (so-called soft law) of international origin, an example of the former being the 1980 Vienna Convention,[5] and the latter being PECL 2000, the UNIDROIT Principles (UPICC) 2004 and 2010 and the DCFR.

UNIDROIT: 4.1.

(1) A contract shall be interpreted according to the common intention of the parties.

(2) If such an intention cannot be established, the contract shall be interpreted according to the meaning that reasonable persons of the same kind as the parties would give to it in the same circumstances.

4.2.

(1) The statements and other conduct of a party shall be interpreted according to that party's intention if the other party knew or could not have been unaware of that intention.

(2) If the preceding paragraph is not applicable, such statements and other conduct shall be interpreted according to the meaning that a reasonable person of the same kind as the other party would give to it in the same circumstances.

DCFR: II. – 8:101 General rules

(1) A contract is to be interpreted according to the common intention of the parties even if this differs from the literal meaning of the words.

[5] United Nations Convention on Contracts for International Sale of Goods (Vienna, 1980) (CISG). https://uncitral.un.org/sites/uncitral.un.org/files/media-documents/uncitral/en/19-09951_e_ebook.pdf, last access on the 4th of August 2022.

(2) If one party intended the contract, or a term or expression used in it, to have a particular meaning, and at the time of the conclusion of the contract the other party was aware, or could reasonably be expected to have been aware, of the first party's intention, the contract is to be
interpreted in the way intended by the first party.
(3) The contract is, however, to be interpreted according to the meaning which a reasonable person would give to it:
(a) if an intention cannot be established under the preceding paragraphs; or
(b) if the question arises with a person, not being a party to the contract or a person who by law has no better rights than such a party, who has reasonably and in good faith relied on the contract's apparent meaning.

II. – 8:102: Relevant matters
(1) In interpreting the contract, regard may be had, in particular, to:
(a) the circumstances in which it was concluded, including the preliminary negotiations;
(b) the conduct of the parties, even subsequent to the conclusion of the contract;
(c) the interpretation which has already been given by the parties to terms or expressions which are the same as, or similar to, those used in the contract and the practices they have established between themselves;
(d) the meaning commonly given to such terms or expressions in the branch of activity concerned and the interpretation such terms or expressions may already have received;
(e) the nature and purpose of the contract;
(f) usages; and
(g) good faith and fair dealing.
(2) In a question with a person, not being a party to the contract or a person such as an assignee who by law has no better rights than such a party, who has reasonably and in good faith relied on the contract's apparent meaning, regard may be had to the circumstances mentioned in sub-paragraphs (a) to (c) above only to the extent that those circumstances were known to, or could reasonably be expected to have been known to, that person.

II. – 8:103: Interpretation against supplier of term or dominant party
(1) Where there is doubt about the meaning of a term not individually negotiated, an interpretation of the term against the party who supplied it is to be preferred.
(2) Where there is doubt about the meaning of any other term, and that term has been established under the dominant influence of one party, an interpretation of the term against that party is to be preferred.

From the point of view of the topic of this book, it seems crucial to understand how the subjective element of interpretation can be related to AI. The subjective method of interpretation directly relates to properties closely related to humans, such as consciousness (awareness) or intention (intent). A contract is entered into by people; even if a legal person is a party to the contract, it always acts through agents

or its organs e.g. the board of directors, which are people. In the case of AI, the decision-making sphere relating to the contract may be, to a greater or lesser extent, transferred to the software. It is hence necessary to analyse how to interpret contracts entered into by AI, when it cannot be attributed with a consciousness.

Several issues need to be considered in this context:

(1) Does the use of AI affect the interpretation of the contract?
(2) Is there any difference between the natural language used by humans and the natural language used by AI, and if so, should this difference be recognised?
(3) Should subjective concepts such as awareness or intent refer to AI or to the party to the contract, and if they pertain to AI, what meaning should be given to them?
(4) Does the party which uses AI have a 'dominant influence' over the party which does not?
(5) Should AI be considered a dominant party?
(6) Should the principle *in dubio contra AI* be recognised in respect of contracts concluded with AI?

A contract is to be interpreted according to the common intention of the parties. Therefore, it must first be determined whether, in interpreting it, it is necessary to have regard to the party of the contract (e.g. a human being) or to the AI who concluded it. In the latter case, it is necessary to determine how to understand the intention of the AI.

When considering the action of an AI in legal transactions, it must be recognised that to understand the content of the contract, neither the intention (and subjective situation) of the contracting party, nor that of the AI itself, perhaps more so, cannot be of exclusive importance. Such an approach remains consistent with the assumption that, except for situations when the AI is a mere conduit, or when it acts within its own legal capability, the AI occupies the position of an agent when entering into a contract. In cases where an agent enters into a contract, it is the intention given by the agent and his level of sophistication that must be relevant for its interpretation; after all, it is the agent's statement that is exclusively visible to the other party and any outside observer. The true impact on the interpretation of the contract by the intentions of the principal and of the agent may vary depending on the scope of the authorisation, i.e. the extent to which the will of the agent himself could influence the content of the contract. If the agent's role was merely to repeat, or possibly to make more precise, the intentions expressed by the principal, it is the principal's intention that must play the decisive role in understanding the party's intention and, consequently, for interpreting the agreement. If, however, the agent shaped the content of the agreement himself, mere reliance on the principal's intention may be insufficient; however, it also remains relevant precisely to the extent that the framework of the authorisation was set.

This is, of course, of direct relevance to understanding the contract entered into by an AI. Let us assume that the user of an autonomous refrigerator specifies in general terms that the AI alone is to purchase food products for him, without the user in any way specifying his preferences; according to the concept of this device, it is to calibrate itself accordingly to the user's expectations, conscious or unconscious,

through contact with the user's environment. If such an AI concludes a contract for the purchase of a specific product (e.g. cheese), an analysis of the user's general intention will provide little in the way of interpreting this contract.

Such an approach to the issue of interpretation makes one wonder how intention could be understood in the case of a contract concluded by AI. Is it at all reasonable to talk about the search for the intention of the parties and a combined interpretation (with a subjective element) of the AI's statement of intent? In the same way that we assessed fault and consent, we also believe the concept of intentional action is needed in the context of contract interpretation, but with some modifications.

It is impossible to look into the mental processes of a human being, much less an AI. In the case of the latter, intention can only be assessed in the light of objective measures resulting from the circumstances in which the contract was concluded. The basic element determining the intention of an AI is its specification, resulting from its registration documents. The purpose of the AI described therein should indicate the intention (objective) of its actions. Thus, in relation to AI, intention (in the subjective sense) boils down primarily, though not entirely, to the purpose or function of a given AI. As we have indicated elsewhere (chapters: consent, abuse of rights) intention is determined by how the other party can understand a particular action. In the context of AI, this must therefore be an understanding based on the publicly-available information on the machine. It must also be linked to the explainability of the action of a particular system; if the purpose of the action of the system is known and the way decisions are made can be explained, then it can be assumed that the intention of the action can be known.

Let us note that such an approach is in fact already present in global jurisprudence. In 2020, a Singapore court questioned a programmer to ascertain the intention of the parties regarding certain transactions in which an AI acted.[6] This type of approach is directionally correct, but obviously far from sufficient. It is impossible to consider that, in the context of contracts entered into by AIs, one should rely on such an uncertain element as the testimony of a programmer. The specification of the performance of the AI, and in fact, everything that the programmer could theoretically say, must be predetermined in the registration parameters, to be available and verifiable.

This approach, it seems, will suffice in most cases. However, unlike human actions, the unpredictability of decisions made by an AI is framed within the narrow confines of a given algorithm. In other words, in the case of a human acting to buy a particular good or obtain a particular service, intention can vary widely and cannot be reduced to any general human 'trait': it is a function of the totality of a person's physical state and psyche, experiences, knowledge, context and so on, which is unique and can continually change. For existing weak AIs, operating within certain specifications, the scope for this freedom, this unpredictability, is very narrow. Let us return to the example of ordering cheese. A human being may have all sorts of

[6] Quoine Pte Ltd vs B2C2 Ltd, [2020] SGCA (I) 02, https://www.sicc.gov.sg/docs/default-source/modules-document/judgments/quoine-pte-ltd-v-b2c2-ltd.pdf, last access on the 4th of August 2022.

whims in this regard; however, the AI will act in an explainable manner based on predetermined parameters; it is these parameters that will suffice to explain AI's 'intention'.

However, we do not claim that the subjective aspect of the interpretation of statements made by AI must always be limited solely to an examination of the parameters of the machine itself.

First, as we signalled above, the subjective elements concerning the one on whose behalf the AI acts may also be relevant, insofar as its will still determines the content of the statement. Indeed, there is no dichotomous division between those statements whose content is determined entirely by humans (with the machine performing only a technical function) and those in which the volitional element is entirely on the side of the machines. Rather, it is a spectrum of the most diverse possibilities for the distribution of decision-making (free will). As AI becomes more advanced, and as a given user increasingly wants to use such modern systems, choice will become more influenced by the AI; consequently, there will be fewer subjective elements on the side of the user to influence the understanding of the content of the statement. In our opinion, this requires a special method of interpretation, or a subcategory of the combined method: it assumes that the subjective element (will, intention of the party) should be, or rather could be, at least in some cases, identified (searched for) in relation to two "subjects" acting autonomously in certain spheres of decision-making. Only the sum of these "wills" or "intentions" constitutes the "intention" of the party to the contract, which can be relevant for the interpretation of the contract. The intention is therefore no longer either the intention of the user or the intention of the AI, it is the resultant of them.[7]

Secondly, a reference to the content of the register will not always be sufficient to understand a party's intention *in concreto*. In the method of combined contract interpretation, what matters in this respect is the consensual intention of the parties. Determining the AI's "intention" therefore never takes place in a manner alienated from the circumstances and intentions of the other party. The intention of the AI would then have to be determined in the same way as that of the other party.

9.3.1 The Conclusion of the Contract by AI as a Circumstance Affecting the Interpretation of the Contract

The circumstances in which a contract was concluded are of special significance in its interpretation.

[7] If someone doubts that such a summary intention has sense, it should be noted that a very similar situation happens when a collective body, consisting of persons with the different power (e.g. the president may dispose two votes), makes a decision.

Is the fact that the contract was entered into by AI a circumstance which is relevant to its interpretation? In our view, this fact is always relevant, and therefore must be considered as part of the circumstances in any case where a contract is interpreted. Obviously, in cases of "ordinary" contracts concluded by humans or their automated agents, this problem does not arise at all; they have traditionally been the only contracts that were concluded, and there was no need to analyze the nature of the entity acting as the party. After all, even in cases where a legal person is a party to a contract, human beings are the causative factor in the process of negotiating, drafting and concluding the agreement, and so are responsible for its shape and content. It is only the action of an autonomous AI that changes this paradigm: a non-human agent appears in the contractual relationship, not merely being a carrier of human will, but acting on its own initiative. Hence, in the case of any contract concluded by an AI, this circumstance alone can potentially affect the understanding of the content and effects of the contract. Consequently, this fact is always legally relevant, and can never be hidden from the person interpreting the contract, primarily the court. The fact that a contract between two subjects is concluded using a quasi-entity may impinge on the understanding of the content of the statements. Even contracts made between humans with the same wording may mean different things and may, potentially, produce different effects to those concluded between a human and an AI. This applies all the more to A2A contracts: the specificity of contracting between AIs should be taken into account as a relevant circumstance for understanding the content of the contract.

The fact that the action of an AI is relevant to the interpretation of the contract may result from a wide variety of issues; the simplest example may be the systemic link of the AI's action to specific markets, or it may be associated with the specific meaning of the words used by the party to the contract in this case, acting through the AI. If, for example, the AI operates in the market of children's toys and enters into a contract that concerns a car, there would be no room for doubt: it would be clear from the very registration parameters defining the purpose of the AI in question that, since it does not operate in the car market, the word "car" had a specific meaning. The intention of the AI cannot be recognized as ambiguous: it should be to conclude contracts in the toy market. Thus, the AI's very specification may, and indeed usually will, impinge on the meaning of the phrases it uses. Of course, it is also necessary to consider the characteristics of natural or legal persons when trying to understand their statement, but this will have greater significance in the scope of AI. Indeed, without linking the AI in question to its scope of action, no meaning can be given to its statement.

9.3.2 The Language of the Contract. In Dubio Contra AI

It is difficult to imagine, that a contract would not be expressed in natural language but, for instance, only in pure code, even if the agreement is A2A. Such a practice may well be recognized as a breach of the requirement for transparency. At least

some translations from the code would be necessary, for example to satisfy auditing procedures. Of course, when the contract is A2H, it is all the more clear that it should be formulated in natural language, even by the means of a model contract. And this language, and the very contract in which it is formulated, needs interpretation.

It may seem that interpreting natural language is nothing special; it is something complicated, but has been easily achieved since antiquity. However, the problem becomes more complicated when negotiations are performed between an AI and a human, or when a the drafter of the contract is an AI. This problem is not one for the distant future, as even today many commercial firms use AI-based chatbots for communicating with clients. And this phenomenon raises a number of difficult questions.

Is the natural language of AI the same as the natural language of humans? Can AI really grasp the meaning of the utterances expressed by humans, or can a human realize that his interlocutor does not speak his "native" language during their conversation? Should the contract drafted by AI in natural language be interpreted in the same way as a contract prepared by a human? The answers are not easy. On the one hand, the natural language given to AI by its creators is assumed to be an imitation of natural human language, but on the other, AI faces many disadvantages, and advantages, when compared to humans in this domain. The capability of AI to understand and process natural language is a matter of Natural Language Processing (NLP). Hendrycks et al. (2021), pp. 2–3 observes that

> Researchers in NLP have investigated a number of tasks within legal NLP. These include legal judgement prediction, legal entity recognition, document classification, legal question answering, and legal summarization (Zhong et al., 2020). Xiao et al. (2015) introduce a large dataset for legal judgement prediction and Duan et al. (2019) introduce a dataset for judicial reading comprehension. However, both are in Chinese, limiting the applicability of these datasets to English speakers. Holzenberger et al. (2020) introduce a dataset for tax law entailment and question answering and Chalkidis et al. (2019) introduce a large dataset of text classification for EU legislation. Kano et al. (2018) evaluate models on multiple tasks for statute law and case law, including information retrieval and entailment/question answering.

It should be noted that this monograph is not a good place to explain or review these works,; it only provides examples of works on the legal NLP and of NLP in general. As such, only some general notices will be made.

Although AI may have a much wider lexicon than even a well-educated human being and have access to vast contextual data bases with legal texts, the issue of legal NLP is beset by numerous problems, and as such, is the focus of many competing concepts and ideas. Hence, the form of NLP method or dataset embedded in an AI system involved in a contract should be taken into consideration during its interpretation (of course, when there is some interpretation problem).

A second possible postulate is justified by the need to protect human parties of the agreements: when an AI is a drafter of the agreement, the rule *in dubio contra proferentem* should be binding. However, it must be considered that this rule has a different justification than in the case where all parties are human or legal persons. When this clause rules the interpretation of contracts of non-Ais, it is because bad faith is suspected or even assumed. If it were to govern the interpretation of contracts

drafted by AI, it is not because of probable dishonesty, but the fact that the interpretatively-doubtful contract was prepared in an ambiguous manner instead of the clearest way possible, as it is obligatory by design. Obviously, this rule should be applied mainly in these situations when a contract is challenged in court.

Even in classical Roman law, the idea was expressed that if the text of an obligation was unclear, it had to be interpreted to the disadvantage of the one who formulated it, as he could, and should, have expressed himself more clearly. From Accurius comes the paremia *in dubio contra proferentem* ("in case of doubt against the declarant"). This paremia is in extensive use in jurisprudence in civil law countries, with the approval of a large part of the doctrine, especially in the field of consumer protection. It emphasizes the increased risk borne by the party that did not formulate the content of the contract; it also stresses the need to protect trust.

> UNIDROIT: Article 4.6 If the contract terms supplied by one party are unclear, an interpretation against that party is preferred.

In this context, there is nothing preventing this principle being applied in regard to AI. So, in view of the repeatedly-mentioned intellectual superiority of machines and the problems with the language used by the AI to communicate with "normal" users, introduced herein, this principle should be extended to all dealings with AI; it should also be recognised that in all cases where AI proposes the content of the contract, not only where the counterparty is a consumer, doubts must be resolved in favour of the other party. This principle should also be extended materially to include cases where the contract is shaped by negotiation between a human and an AI (H2A). In addition, as such contracts may potentially be "tainted" by intellectual-emotional imbalance, it would be reasonable for any ambiguity to be clarified in favour of the human party, and never the AI. Therefore, we consider it justified to assume *in dubio contra AI* as the rule for interpretation. This rule should apply to both professional and general trading in cases where both parties have agreed on the content of the contract. Let us take an example. Assume that two parties are negotiating a contract and one of them uses an AI: in this case, the latter should be considered dominant from the point of view of interpretation, and doubts should be interpreted in favour of the other party. Of course, this rule cannot apply if it is the other party who submits a model contract that the AI merely accepts.

The third postulate regarding interpretation, which may serve as an exception from the *in dubio contra AI* rule in certain situations, is connected to the fact that the language AI uses may be determined by registration parameters. So, if it follows from the registration information of the AI that it operates in a certain field, for a certain purpose, then its use of certain terms should be interpreted in accordance with those registration elements rather than *contra AI*. It should be noted that the concept of AI operation in legal transactions presented herein is based on the fact that the other party must have full access to these registration elements, and therefore knows what to expect. If, in the light of certain parameters, AI can and does act only in such a way that its statement can be translated in only one particular way, it should not be possible to come to another interpretation, even if the circumstance could theoretically be translated differently. This apparent collision of meanings can be reconciled

by recognizing that when viewed through the prism of registration, the statement must be considered clear and unambiguous. Only if the analysis referring to the registration data did not yield the interpretation decision i.e. the statement still could be interpreted in different ways, could one go to the *contra AI* rule.

9.3.3 AI as a Dominant Player in the Negotiations

When considering how the fact that one of the parties is represented by AI should influence the interpretation of the contract, it should first be examined whether such a party may, for the mere reason of using AI, be regarded as acting with dominant influence. For reasons relating to the need to protect humans as the weaker party (as justified in Chap. 10), the presumption that a party acting with AI is the party with dominant influence, applied in the moment of contract interpretation, may be considered. Since the use of AI in trading is justified by its greater effectiveness and intellectual superiority over a human being in that field. Artificial Intelligence systems act without emotions and with mathematical precision, and their efficiency is determined in advance and remains constant through their use, at least in certain parameters. Therefore, if the other party is not an AI, it is highly likely that it will have inferior intellectual capabilities within the field in which the AI operates, than the AI.

Of course, in cases where an AI is the party of an agreement, or where the party uses an AI agent, does not necessarily bestow an advantage that should be classified as a dominant influence. It cannot therefore be said that AI systems always have an advantage and should always be recognized this way. We believe, however, that it seems reasonable to call for the introduction of certain presumptions in this regard: assuming that the party using an AI should be deemed to have a dominant influence when interpreting a contract does not exclude proving the contrary. Therefore, the party who uses an AI should be required to prove that its position was equal to, or worse than, that of the human party. Only such a distribution of the burden of proof would remain in line with the general ethical demands related to the regulation of AI; this is the only way to guarantee the protection of the other party against making use, even in a very subtle way, of the advantage offered by the use of AI, and to protect human autonomy to the fullest extent possible. At the same time, the construction of the rebuttable presumption, i.e. that AI has a dominant influence, is a proportionate instrument for protecting this autonomy: it does not unduly interfere with the freedom of contract, nor does it grant the AI counterparty an excessive, unjustified privileged position.

However, following the introduction of such a presumption, it would be necessary to develop conflict rules for situations in which existing rules indicate the other party to be the dominant party. It would therefore be a question of determining which advantage is more important: the advantage stemming from the use of AI or the advantage resulting from, for example, the professional activity carried out by the other party. It may be difficult to resolve the conflict between these presumptions on

a general level, due to the wide potential diversity of potential factual situations. In the case that a contract is concluded between a professional trader and a consumer using AI, the problem is of a broader significance and will relate not only to the issue of contract interpretation. With the development of laws protecting humans from the risks associated with the use of AI, collisions with the consumer protection system may increasingly arise. Taking the example of an autonomous fridge placing orders in a small vegetable shop, the user of the fridge, as a consumer, has special protection in his dealings with the seller, as will the shop due to the fact that the contractor uses AI. It is possible that the legal preferences of both parties will be mutually exclusive, in which case conflict mechanisms will have to be created.

In the case of the use of AI, which is not registered, it should consequently be considered that such action does not fall within the permissible contractual framework; consequently, a party may use all the tools available to protect itself. For instance, it may demand the annulment of the contract. However, if a party does not decide to annul the contract and the problem of its interpretation arises, it should be considered that *a minori ad maius* the other party is also dominant in this case. Since even an AI that is registered, and therefore controlled and supervised by the state, is still dominant, it is all the more necessary to also consider an AI that is not subject to any control, or inadequate or insufficient control, as dominant.

9.4 Due Diligence in Contracts Involving AI

As the problem of due diligence is analysed in Sect. 11.5, we can only add some additional remarks.

In contractual relations, the issue of due diligence obviously has a much broader significance than just in terms of establishing good faith or liability for damages. The debtor must perform the obligation in a diligent manner; a lack of due diligence becomes the basis for attributing contractual fault to the debtor. How is the requirement of AI for due diligence defined in relation to a contractual relationship? Should the level of diligence be lower or higher as in the case of a human being (more broadly: another entity) performing an obligation, or should it remain the same? The answer to this question must take into account the following elements:

First, the level of expected diligence must be related to the performance specification of the AI system in question. It cannot be abstractly assumed that the level of diligence remains the same, regardless of the type of system and the objectives that are set before it. In relation to humans, one traditionally builds a model of a diligent debtor (or a diligent professional in a given field of trade) and compares actual behaviour with the hypothetical behaviour of such an imagined model. While it is possible to create a single model of a diligent person, such as a diligent merchant or diligent debtor, this does not seem possible for AI; no single type of AI exists, and their diversity makes it difficult even to find a lowest common denominator around which it would be possible to build a model of behaviour. Furthermore, it is important to remember that most AI systems operate with a certain percentage

level of effectiveness. If this level of performance is indicated in the registration specification, then it does not seem possible to demand that a given AI performs more diligently than is indicated by this level. For example, if a given system is 90 per cent accurate in diagnosing a particular disease, then the level expected of this AI is 90 per cent. As such, this is the specification of a given AI that can determine the level of care required. However, this is not the only parameter determining the level of due diligence demanded or required.

Secondly, to prevent a person who does not use an AI from being treated more rigorously by the law than one who does, the level of due diligence required for an AI cannot be less than that of a human being. Therefore, if we demand a certain level of due diligence from a party, the fact that a person uses AI for concluding or performing a contract cannot entail a reduction in the required diligence.

Thirdly, notwithstanding the above, it must be accepted that the required level of diligence should be determined in relation to the current level expected of AI systems. Thus, for example, if the standard is a 95% success rate of some price realization, this would be the required level of diligence for this type of AI system.

9.5 Performance of Contracts by AI

Regardless of whether a contract has been concluded using AI, intelligent systems can be used to perform such a contract. It should be noted at the outset that two fundamentally different types of contracts can be distinguished from this point of view:

(1) those in which the performance of a party (or parties) takes the form of a declaration of intent and does not require any factual action (so that the performance of the contract is constituted only by a legal act, like its conclusion); this may be seen in the case of the sale of shares on the stock exchange or the conclusion of a contract in execution of a prior preliminary contract, and
(2) those that, at least in part, require specific factual steps to be taken (e.g. the provision of hotel services, the sale of goods to be delivered to the buyer). This group may also include such activities that are performed by a robot, i.e. those that are already directly related to the performance of some physical work such as loading goods, performing construction work or manufacturing items in a factory. However, they can also be activities that only take place in the digital world, such as translation of text, or performing medical diagnoses based on data analysis.

In both of these conventionally-delimited groups, the typology here is of course only illustrative, neither strict nor dichotomous, an AI may be used at the stage of execution, albeit to various extents. However, the way AI is used, and any consequent problems for civil law, differ from one market segment to another.

Regarding the first group, these represent contracts whose performance will require the submission of a declaration of intent, the content of which will be

determined by AI; hence, it will not be just a system of automatic performance or self-performance, unlike smart contracts. In this case, therefore, the same structural problems arise regarding the attributability of the AI's statement to the user as seen in the case of contracting (Chap. 5). An AI making a declaration of intent to perform a contract should therefore be regarded as an agent, acting on behalf of and with effects for the contracting party. The consequences of the action of an illegal AI, acting outside the scope of its authorisation, or where there are defects in the declaration of intent, should therefore be assessed in the same way as in the case of the conclusion of contracts (cf. Chap. 5).

In the second case, where the performance of the contract will consist of specific factual acts by the AI, it is crucial to determine whether this is an act of the contracting party itself (as in the case of non-autonomous machines) or whether it should rather be considered an act of a third party. At least in some cases, such a distinction may be relevant to the assignability of liability, and even the admissibility of action by the AI itself.

The rules for the performance of contracts involving a third party, and the debtor's liability for such actions, are quite similar between the various European legal systems. By way of example, we may refer to the DCFR:

III. – 1:102: Definitions

(2) Performance of an obligation is the doing by the debtor of what is to be

done under the obligation or the not doing by the debtor of what is not

to be done.

III. – 2:106: Performance entrusted to another

A debtor who entrusts performance of an obligation to another person remains responsible for performance.

III. – 2:107: Performance by a third person

(1) Where personal performance by the debtor is not required by the terms regulating the obligation, the creditor cannot refuse performance by a third person if:
   (a) the third person acts with the assent of the debtor; or
   (b) the third person has a legitimate interest in performing and the debtor has failed to perform or it is clear that the debtor will not perform at the time performance is due.

(2) Performance by a third person in accordance with paragraph (1) discharges the debtor except to the extent that the third person takes over the creditor's right by assignment or subrogation.

(3) Where personal performance by the debtor is not required and the creditor accepts performance of the debtor's obligation by a third party in circumstances not covered by paragraph (1) the debtor is discharged but the creditor is liable to the debtor for any loss caused by that acceptance.

When performing an obligation, the debtor either acts alone or with the help of other persons (assistants, executors); the difference between these situations is of a normative nature and boils down to determining whether the person with whose assistance the debtor performs the obligation is a separate entity (natural or legal person). If he is not, because he helps himself with a machine or algorithm, his action is treated as that of the debtor's tool, and thus the action itself is attributed directly to the debtor; it is simply his action. It should be noted, however, that for a long time, any tools which have autonomy, i.e. animals, have been treated differently. If a person uses an animal, liability is shaped in a different way than when using an inanimate tool, such as a hammer or a more advanced machine. When considering the action of an AI performing an obligation, there are therefore at least three possibilities for the model qualification of this action: a third party, an autonomous tool (like an animal) and a passive tool.

In the case of the use of AI, whose role is to make choices and decisions beyond the debtor's reach, the matter is undoubtedly complicated. Imagine a contract to provide a navigation service with a legal person: it seems natural to consider the AI navigation system as a tool through which this person performs its obligation. However, such a qualification will not be so obvious if we consider its counterfactual aspect. The person who uses AI to perform an obligation in fact delegates part of his decision-making capacity, necessary for the performance of the contract, to another entity. Since it is not that person only who decides on the manner of performance. Taking the simple example of car navigation: the provider of this service does not determine the route; it is "created" by the machine.

It seems that a hybrid solution should be sought in relation to AI: AI is both a tool for performing an obligation and a third party (or quasi-person) with whose help the obligation is performed. To take another example: x has committed to guard a house. In order to perform this obligation, he has trained two dogs to guard the premises, at a time when he himself does not directly patrol the premises. Is the action of the dogs the action of the debtor himself or rather of a "third party"? The statement that the animal is only a tool in the hands of the person is counterfactual in this case, because there is no certainty as to how the dogs would concretely behave; at the same time, it is impossible to speak of the dogs as a "third party" *sensu stricto*, with whose help the debtor performs the obligation, because, as we have already described, animals are not persons.

This problem becomes glaringly apparent in the case of AI, since the performance of an obligation is based on the delegation of authority to act to fulfil it. Such a qualification is easier for AI than for animals. By performing the contract, the AI obtains a normative argument for its eventual subjectivity. However, it does not have to have the full legal subjectivity; the narrow, punctilious extent is enough to hook its qualification as a third party.

We previously established that when using an AI, the distribution of decision-making between the user and the AI system itself may vary depending on its sophistication, scope and purpose of action. In other words, the will for a particular action may be shared very differently between AI and the user depending on the situation.

Similarly, the problems concerning the role of AI, i.e. as a tool or as a third party used by the debtor in contractual relations, are complicated by a whole range of possible factual states: it is not possible to reach such a dichotomy. Instead, we can speak of a whole spectrum of situations, i.e. a greater or lesser empowerment or, on the other hand, objectification (un-tooling) of AI. In those cases where autonomy is essential for the performance of the contract, we may assume that the AI acts as a third party. In other cases, however, when the essential elements of the contract are performed by humans, or more precisely, when the parameters of this performance that are essential are determined by humans, then the AI can be regarded as only a tool.

Returning to the example of an autonomous refrigerator, it could be regarded as playing the role of the user's agent in any sales contract which it concludes; it acts on his behalf and on his account. We can assume that it uses his money, which may be kept in a separate fund belonging to the AI, or be taken directly from the user's account, which is solely at the disposal of the AI; this is discussed in more detail in Chap. 8. The AI then performs the obligation of the concluded contract, i.e. fulfils its performance by transferring the funds to the seller's account. The user, of course, has no influence on the conclusion of a particular contract. Due to the autonomous nature of the AI action under this contract, its action in this respect should be qualified as an independent performance by a third party, not as that of a tool. However, the action of the AI, as a third party, is equivalent to the performance of the debtor itself, i.e. the user of the refrigerator. The correctness of the action of this person is the responsibility of the debtor. As a rule, the other party may not refuse to accept the performance provided by a third party in this sense, and may not demand payment directly from the user, but may refuse to accept the performance autonomously initiated by the fridge. This will be the case not only when the performance is in the form of money.

Of course, the qualification of the AI as a third party to whom the debtor entrusts the performance of the obligation, and not merely a tool in his hands, does not weaken the position of the creditor, since, undoubtedly, the debtor is liable for both the action of this helper and his own action (DCFR III: 2:106 cited above). What is more important, however, is that regarding the subjective position of the AI as qualitatively different from that of a mere tool may strengthen the position of the other party to the contract (the creditor).

The significance of this problem will be seen where the debtor is to act personally in performing the obligation. As a rule, the debtor only has to perform an obligation personally if the situation fulfils certain conditions, for instance when required by the content of a legal act, from the law or from the nature of the performance. In BGB, it is formulated in the following way:

Section 267
Performance by third parties

(1) If the obligor need not perform in person, then a third party may also render performance. The consent of the obligor is not required.
(2) The obligee may reject the performance if the obligor objects.

Another example of similar regulation may be PECL:

Article 7:106: Performance by a Third Person
(1) Except where the contract requires personal performance the obligee cannot refuse performance by a third person if:
  (a) the third person acts with the assent of the obligor; or
  (b) the third person has a legitimate interest in performance and the obligor has failed to perform or it is clear that it will not perform at the time performance is due.
(2) Performance by the third person in accordance with paragraph (1) discharges the obligor.

In a standard situation, it will therefore not matter whether the debtor used certain tools or whether he commissioned another person to perform the contract; it will therefore also be irrelevant whether the use of AI should be classified as an action of a third party. However, in some cases, it is necessary to perform the contract personally. Can a debtor who is to perform an obligation personally use AI, and if so, can he do so to a certain extent? Let us take the following examples: can a babysitter hired by the parents to personally take care of their child then deploy an AI to talk to her charge, or can an artist from whom someone has commissioned a portrait use AI to create it? Of course, the contract itself may stipulate whether and to what extent it is permissible. Alternatively, the possibility of using AI may be implicitly accepted by the parties, at least to some extent, given the circumstances; e.g. it is clear that a caregiver can turn on AI to monitor a child's sleep, or an artist who is known to use new technologies supports his works with the use of AI. However, if the contract does not mention this and it is not apparent from the circumstances, and the contract is to be performed personally, does the use of an AI fall within the scope of the "personal" performance of the contract?

Assume a publishing house orders a new crime novel from a well-known writer. Would proper performance of the contract involve the writer pressing "Enter" and activating the AI to write the book for him? One can have serious doubts in this respect; although, it should be noted that in the field of art, the "brand" of a creator associated with a work is often more important than who actually created it, as evidenced by the history of art from Rubens to Damien Hirst. The doubts that arise in this case boil down to the problem of defining the action of a third party in performing an obligation by the debtor. Where the personal action of the debtor becomes an essential feature, the use of an AI becomes problematic. This applies at least to those cases where the role of the AI concerns the essence of the debtor's main performance. In the example of the writer given above, there seems a fundamental difference between using AI to write a book and using AI, for example, to correct errors, improve style or find gaps or inaccuracies in the text. Keeping with the

assumption that autonomy is the key element in the operation of AI, it should be considered that in those cases where the performance of the debtor is in some way and in a broad sense related to decision-making, and where the debtor is supposed to act personally, the delegation of these decisions to AI (in other words: authorizing AI to take these actions) must, in the context of the performance of contracts, be regarded as the action of a third party. Since in the light of our findings in Chap. 7, the AI should be regarded as an author.

In this context, it is also crucial to determine what is meant by "relevant action" and identify the "essential elements" of the contract being performed. When discussing the effects of an autonomous AI, it should be considered that what is at stake are those elements which are decisively important from the point of view of the creditor's interest. If the creditor demands personal performance, it is obvious that he is concerned with the personal performance of certain key elements of the contract, and not with auxiliary elements that do not determine the essence of the performance. Whether these are key (objectively) elements of the contract depends on the context. To sum up, accepting the qualification of AI as a "third party" and not merely a "tool" seems to strengthen the position of the creditor who, in the case of at least some performances, has more influence on whether, and if so to what extent, AI can be used.

9.6 Information Obligation

In any situation where a contract is performed by AI, the question arises whether the other party to the contract should be informed about it. We have established (Chap. 5) that the use of AI at the conclusion of the contract must be publicly acknowledged. The same *rationale* argues that adequate information about the operation of AI, concretized and verifiable in a public register, must also appear at the stage of contract performance, on pain of being deemed to have performed the contract improperly. In contrast, the debtor, as a rule, does not have to inform the creditor that he will use third parties in the performance of the obligation. The question arises, therefore, whether the fact that this third party is an AI is so important that it affects the debtor's obligations. In our view, it does.[8] The performance of an obligation using intelligent systems introduces a significant non-human factor into the entire obligation relationship. To delegate authority to AI is to go beyond the traditional framework in which the obligation was performed by humans using tools that convey their will. Besides, the use of AI while performing a contract associated with processing personal data or making decisions about human affairs

[8] When AI is used in medicine (not only in the performance of medical contracts), such an information obligation appears self-evident and conditions the effectiveness of the patient's consent, in more detail: Pfeifer-Chomiczewska (2022).

may be recognized as an infringement of fundamental rights, this actually existing and those postulated. Here it should be noted that:

> **Article 22 GDPR Automated individual decision-making, including profiling**
> The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
>
> Paragraph 1 shall not apply if the decision:
>
> (1) is necessary for entering into, or performance of, a contract between the data subject and a data controller;
> (2) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
> (3) is based on the data subject's explicit consent.

And the new fundamental right postulated by Ferdinand von Schirach:[9]

> **Article 3 - Artificial intelligence**
> Everyone has the right to know that any algorithms imposed on them are transparent, verifiable and fair. Major decisions must be taken by a human being.

It is of course more difficult to determine at what point such information should appear, what the possible consequences of a lack of such information might be and what rights a creditor who has been informed of the use of AI would have. As a general rule, based on civil law, the creditor cannot object to the use of third parties to perform the obligation, except in the case of obligations to be performed in person, and in such cases, the debtor does not even have to inform the creditor of the use of other persons. Of course, the creditor, unless otherwise agreed, also cannot oppose the debtor's use of certain tools, assuming that their use does not violate the law or certain customs or rules. Given this, it is difficult to see how a creditor could object to a debtor's use of AI in the performance of the contract.

In our view, however, the question of informing about the debtor's use of AI in performing an obligation should be assessed differently. It should not be regarded as excessive to require the debtor inform the creditor prior to using AI, although not necessarily prior to the conclusion of the contract.

However, the assumption that the debtor will have to be informed about the use of AI needs to be clarified. Several models can be imagined for such a situation.

The most far-reaching model would involve assuming that the debtor has to be informed even before the conclusion of the contract. This would lead, firstly, to the creditor being fully informed of the circumstances and being able to take appropriate measures (e.g. withdrawal from the agreement). Secondly, it would lead to the

[9] https://www.jeder-mensch.eu/informationen/?lang=en.

conclusion that the lack of such information entitles the other party to withdraw from the contract, or alternatively, that such performance of the contract is deemed to be improper. Thirdly, if the debtor only decided to use AI after the conclusion of the contract, he would have to obtain the creditor's subsequent consent. This model seems to be one to consider in a consumer context.

By contrast, a softer model would assume that the information must be given before the contract itself is performed, but not necessarily before the contract is concluded, with the creditor not being able to object to the use of AI or rescind the contract in such a case. Although the use of AI would not therefore be accompanied by any potential sanction, providing such information to the creditor would possibly furnish him with certain tools (evidentiary facilities) in the event of improper performance. In contrast, failure to provide such information could be considered as wrongful performance, and the range of sanctions in this case could also vary from damages to the possibility of withdrawal from the contract.

In formulating conclusions in this respect, it is important to bear in mind, on the one hand, the diversity of AI and how it is used in the performance of a contract, and on the other hand, that the use of AI in number of contracts is likely to almost become standard. Let us assume that someone enters into a contract to create an architectural design or to construct a specific building. It is quite likely that, to a greater or lesser extent, 'ordinary' tools (including IT) will be supplemented with AI to speed up the execution of the contract and improve the results. The use of AI in contract performance may be primary or it may be entirely tertiary (e.g. the use of car navigation by a driver working on a construction site). In terms of information, the experience of the GDPR can be successfully used here: a debtor who intends to use AI for the performance of a contract shall inform the creditor, whereby the scope of information would have to include the AI's "identity" and basic parameters as well as registration data; this would allow the legality of the AI to be verified and to provide full knowledge about it. However, it is inconceivable that the creditor could prohibit the use of these tools or consider their use as grounds for withdrawal. Such a general rule cannot be introduced for purely practical reasons alone, nor does it seem necessary. A (blocking) power of this kind could only be introduced if it were deemed to be justified by the objectively-assessed interest of the creditor. However, at least when assessing the case *in abstracto*, no such connection can be discerned.

In contrast, the issue of performance of specific contracts may be different. In this case, due to the specific nature of the obligation, the fact that the debtor makes use of AI to some specific extent may no longer be neutral for the debtor. The issue thus resembles the problem associated with the debtor's personal action; however, it has a broader significance. It is no longer just a matter of whether the debtor performs personally but may also be a matter of whether the obligation is performed by humans or, to some extent, by other autonomous entities. As a matter of principle, a similar rule should be adopted as in the case of personal performance by the debtor: as a general rule, the debtor can also perform the obligation by means of autonomous entities unless otherwise agreed or unless the nature of the obligation requires otherwise. Consequently, a creditor's demand that the debtor not use AI would be unjustified, unless otherwise expressly stipulated in the contract. Of course, just as it

is possible to stipulate in the contract that the obligation is to be performed personally by the debtor, it is also possible to stipulate that it cannot be performed using an AI.

With regard to the consumer, we considered it worthwhile to adopt a model in which it is necessary to make it clear to all parties that the performance of the contract will take place using AI to some extent, even before its conclusion. This obligation would be independent of the obligation to inform that the conclusion of the contract itself takes place using AI. However, this opens up a further question as to whether such informed use of an AI itself should always be legally transparent, or whether a stipulation that the debtor will use an AI may not be an illicit contractual term in consumer relations. The concept of prohibited contract terms is so flexible and capacious (see the Consumer Directive) that it not possible to exclude *in concreto* such a qualification, although the mere reservation does not yet imply the shaping of the consumer's rights or obligations. This could therefore apply to those situations where the use of AI will adversely affect rights and obligations. This could be the case for those relationships where an element of personal performance is commonly accepted, such as for contracts concerning the performance of medical or care services. In such a case, the use of AI will be a significant factor affecting the proper performance of the contract. With the development of care robots, a clause in the model contract provided by a care home that it will use medical robots may, at least in certain situations, be considered prohibited, especially if the scope of this entrustment is not specified.

References

Books and Articles

Chalkidis I, Fergadiotis M, Malakasiotis P, Androutsopoulos I (2019) Large-scale multi-label text classification on EU legislation. Proceedings of the 57 Annual Meeting of the Association for Computational Linguistics. 6314–6322. https://aclanthology.org/volumes/P19-1/, last access on the 4th of August 2022

Duan X, Wang B, Wang Z, Ma W, Cui Y, Wu D, Wang S, Liu T, Huo T, Hu Z, Wang H, Liu Z (2019) CJRC: a reliable human-annotated benchmark dataset for Chinese judicial reading comprehension. ArXiv, abs/1912.09156. https://arxiv.org/pdf/1912.09156.pdf, last access on the 4th of August 2022

Durovic M, Janssen A (2019) Formation of smart contracts under contract law. In: DiMatteo LA, Cannarsa M, Poncibo C (eds) The Cambridge handbook of smart contracts, blockchain technology and digital platforms. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781108592239.004

Hendrycks D, Burns C, Chen A, Ball S (2021) CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review. 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks. https://arxiv.org/pdf/2103.06268.pdf, last access on the 4th of August 2022

Holzenberger N, Blair-Stanek A, Van Durme B (2020) A dataset for statutory reasoning in tax law entailment and question answering. In NLLP@KDD, 2020. https://arxiv.org/pdf/2005.05257.pdf, last access on the 4th of August 2022

Kano Y, Kim M, Yoshioka M, Lu Y, Rabelo J, Kiyota N, Goebel R, Satoh K (2018) COLIEE-2018: Evaluation of the Competition on Legal Information Extraction and Entailment. JSAI-isAI Workshops. https://papersdb.cs.ualberta.ca/~papersdb/uploaded_files/1646/paper_Kano2019_Chapter_COLIEE-2018EvaluationOfTheComp.pdf, last access on the 4th of August 2022

Pfeifer-Chomiczewska K (2022) Intelligent service robots for elderly or disabled people and human dignity: legal point of view. AI Soc. https://doi.org/10.1007/s00146-022-01477-0

Woebbeking MK (2019) The Impact of Smart Contracts on Traditional Concepts of Contract Law. JIPITEC 10:106. https://www.jipitec.eu/issues/jipitec-10-1-2019/4880/JIPITEC_10_1_2019_106_Woebbeking, last access on the 4th of August 2022

Xiao T, Xia T, Yang Y, Huang Ch, Wang X (2015) Learning from massive noisy labeled data for image classification. CVPR. http://www.ee.cuhk.edu.hk/%7Exgwang/papers/xiaoXYHWcvpr15.pdf, last access on the 4th f August 2022

Zhong H, Xiao C, Tu C, Zhang T, Liu Z, Sun M (2020) How does NLP benefit legal system: A summary of legal artificial intelligence. ArXiv, abs/2004.12158, https://arxiv.org/pdf/2004.12158.pdf, last access on the 4th of August 2022

Documents

Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council, L 304/64, https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=celex%3A32011L0083, last access on the 30th of October 2022

United Nations Convention on Contracts for International Sale of Goods (Vienna, 1980) (CISG). https://uncitral.un.org/sites/uncitral.un.org/files/media-documents/uncitral/en/19-09951_e_ebook.pdf, last access on the 4th of August 2022

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), L 178/1, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031, last access on the 4th of August 2022

French Civil Code 2016, the English translation provided by J. Cartwright & Bénédicte Fauvarque-Cosson, http://translex.uni-koeln.de/601101/_/french-civil-code-2016/, last access on the 4th of August 2022

Chapter 10
Abuse of Right

10.1 Introduction

The concept of abuse of rights can be found in various forms of legislation and branches of law (e.g. forum shopping, abuse of process, malicious prosecution, tax avoidance). Here, however, we will examine the concept in the context of civil law, and primarily from a European perspective. In general, we can say that the term *abuse of rights* refers to those situations in which the entitled subject acts within the limits resulting from legal regulations and contractual relations, but in a way which is contrary to the aim of the law, or one that serves to cause harm to the other party or cannot be accepted for axiological reasons. In some countries, the concept is only present in case law or doctrine, while in others it is directly mentioned in the text of the law. In some systems, the source of the concept is derived from fault, while in others, it is a concept of good faith.

For example, in Belgium, the grounds for attributing the abuse of rights, related to the concept of fault, is stated in the Civil Code:[1]

> Art. 1382: Tout fait quelconque de l'homme, qui cause à autrui un dommage, oblige celui par la faute duquel il est arrivé, à le réparer. [Any act of man, which causes damage to another, obliges the person by whose fault it occurred, to repair it.]
>
> Art. 1383: Chacun est responsable du dommage qu'il a causé non seulement par son fait, mais encore par sa négligence ou par son imprudence. [Everyone is responsible for the damage he has caused not only by his own act, but also by his negligence or imprudence.][2]

In France such grounds are related to the concept of good faith. The French Civil Code (in the version valid from 17 February 1804 till 01.10.2016, https://www.

[1] Léonard (2016).

[2] https://www.ejustice.just.fgov.be/img_l/pdf/1804/03/21/1804032153_F.pdf, last access on the 4th of August 2022.

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_10

legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006070721/1804-02-17/, last access on the 4th of August 2022) indicates

> Art. 1134 (3) Les conventions légalement formées [. . .] doivent être exécutées de bonne foi. [Agreements legally concluded [. . .] must be performed in good faith.][3]

While the Swiss Civil Code directly use the term "abuse of right" and refers to the concept of good faith:

> Chacun est tenu d'exercer ses droits et d'exécuter ses obligations selon les règles de la bonne foi. L'abus manifeste d'un droit n'est pas protégé par la loi.
>
> [Everyone is expected to exercise their rights and fulfil their obligations according to the rules of good faith. The manifest abuse of a right is not protected by statute.]

And § 226 BGB states:

> Die Ausübung eines Rechts ist unzulässig, wenn sie nur den Zweck haben kann, einem anderen Schaden zuzufügen. [The exercise of a right is unlawful if its purpose is only to cause harm to another.]

The UNIDROIT (Article 1.7.) regards abuse of right as the most characteristic example of bad faith. The document explains:

> A typical example of behaviour contrary to the principle of good faith and fair dealing is what in some legal systems is known as "abuse of rights". It is characterised by a party's malicious behaviour which occurs for instance when a party exercises a right merely to damage the other party or for a purpose other than the one for which it had been granted, or when the exercise of a right is disproportionate to the originally intended result.

From a legal perspective, a significant role in the abuse of right by the actions of an entity is played by extra-legal normative systems, encoded in such terms as morality, decency, equity or good faith. In other words, it is not only the objective fact of what happened that is important, but also the subjective circumstances, such as awareness, will, psychological attitude and so on. Based on this understanding, the concept of abuse of rights is clearly adequate for evaluating any action taken by and AI.

We should start from the observation that in the case of a human using an AI intentionally and as a tool for obtaining some prohibited or ethically-questionable result, it is quite obvious that this represents an abuse of rights. In such a case, however, it is the behaviour of the operator that will be assessed, not that of the AI itself.

It is also undisputed that in certain contexts, the very use of AI and in particular, the fact that AI may be awarded some form of legal subjectivity, may *in concreto* be considered an abuse of this legal construct. This situation would be analogous to that of the abuse of the legal person construct, when it is used knowingly to defraud creditors. In both cases, the problem should be assessed at the level of behaviour of the entity using the AI, or indeed any other tool or normative construction, to achieve prohibited or axiologically-questionable goals; it would also be appropriate for an

[3] https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006070721/LEGISCTA000006150240/#LEGISCTA000006150240, last access on the 4th of August 2022.

entity that aims in this way to improve its position unlawfully (e.g. by avoidance of liability).

The new problem, however, concerns the assessment of the actions of the AI itself as potentially constituting an abuse of rights, rather than the actions of the operator. Where the action is performed by the AI, this raises the need for a new perspective on these problems: it is necessary to determine whether subjective circumstances that could be attributed to a human can also be applied to the performance of an autonomous system, and if so, describe how this can be the case. It is therefore necessary to consider:

1. Is it possible to regard AI as abusive if, in relation to a human being, such a classification would require the adoption of certain subjective elements?
2. Is it possible to regard AI as abusive for reasons other than those recognised in relation to human action, in relation to its intellectual advantage over humans?
3. Can the machine-human relationship give rise to the need for a new definition of abuse of rights, and whether new, previously unknown forms of abuse of rights may emerge that do not fit into the existing framework of this concept?

As a matter of principle, any potential to act in a way that would constitute a typical abuse of the law should be eliminated at the design stage of the AI, and later on during its testing and registration, just as acting in an unlawful or circumventing manner should be. The purpose of the AI must be clearly described and strictly regulated; it should not be technologically permissible to go beyond these limits. In other words, the rules relating to AI must require that, on the technological side, the AI should not be able to exercise the law for any purpose other than its intended purpose. Any transgression beyond this purpose must be considered unlawful.

It is, of course, more difficult to identify any problem of legal abuse which would be manifested in concrete ethical violations at the level of design and registration. The discourse relating to AI design largely revolves around this very issue: how to shape an ethical AI. It is not our task to describe this discussion—let us take it for granted that these demands will be met. AI HLEG ETHICS 2019 points to four ethical principles which must be respected in order to ensure that AI systems are developed, deployed and used in a trustworthy manner: (1) Respect for human autonomy; (2) Prevention of harm; (3) Fairness; (4) Explicability. There is no doubt that AI systems should act ethically, and therefore, theoretically, there should be no risk that they may exercise the law in an ethically unacceptable manner. The potential for the machine to employ unethical action as its *modus operandi* must be eliminated at the design level. Therefore, if a system were to operate in such a way, it would have to be considered unacceptable, i.e. contrary to the principles of permissibility. This issue is approached in a similar way in Proposal 2021.

However, it does not seem possible to adopt such a principle in a consistent manner, i.e. one that assumes that any approval for use (i.e. registration) excludes unethical actions. After all, it must be borne in mind that unethical action is an evaluative, intangible category that cannot be easily written into code. Even with regard to high-risk AI, only actions that will result in the "elimination or reduction of risks as far as possible through adequate design and development" are proposed

(Proposal 2021 Article 9.4a) and it is not demanded that AI is constructed in such a way that the risk does not exist at all, which logically speaking, arises from the definition of a high-risk AI.

For an AI operating in civil law transactions, especially at the contractual level, it is indeed impossible to expect that a machine cannot act in such a way that could be qualified as an abuse of rights; this follows from the obvious observation that such an assessment is made *post factum*, taking into account all the circumstances of the case. Rather, it must be flexibly assumed that the registration system should, as far as possible, prevent the operation of systems which will not meet ethical requirements; however, this does not exclude an *in casu* examination as to whether the system in question has breached those rules. Such examination is necessary because these principles are simply variable, evaluable and, for the most part, intangible, and can only be made concrete against the background of a unique factual situation. Moreover, ethical or axiological standards change over time: even the best-designed AI may not be sufficiently progressive in terms of current ethical trends, especially since the assessment of a particular behaviour will take place after the event. Consequently, even the best intentions of AI developers may not be enough to create a machine that fully meets the assumed ethical requirements. This may raise specific civil law problems.

Consequently, it must be considered that the action of a certified AI can also be considered as contrary to axiological principles, even if no one intended to act unethically: neither the creators of the AI, the user, nor the AI itself.

10.2 Abuse of Rights in the Context of the Principle of Respect for Human Autonomy

AI HLEG ETHICS 2019 characterizes the principle of respect of human autonomy in the following way:

> The fundamental rights upon which the EU is founded are directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems. (p. 12).

While the *European Civil Law Rules in Robotics* (2016), p. 21 expresses the concept of autonomy of a human confronted with an AI more precisely and in a way closer to the legal perspective:

> [W]e need to establish a general principle that the robot should respect a person's decision-making autonomy. This would then mean that a human being should always be able to oblige a robot to obey their orders. Since this principle could pose certain risks, particularly

> in terms of safety, it should be tied to a number of preliminary precautions. Where the order received could endanger the user or third parties, the robot should, first of all, issue a risk alert that the person can understand. This would mean adapting the message depending on the person's age, and level of awareness and understanding. Second, the robot should have the right to an absolute veto where third parties could be in danger. As such, the robot could not, then, be used as a weapon, in accordance with the wishes expressed in the motion for a resolution in the paragraph on the "licence for users", which asserts that "you are not permitted to modify any robot to enable it to function as a weapon".

However, the above postulates may face legal and factual obstacles that will be difficult to overcome. The legal problem is revealed in the light of the fact that that maintaining human autonomy against AI will mean taking responsibility for its action. Given the high standard of action performed by an AI in a given field, preserving human autonomy must entail a *de facto* lowering of the standard practised in that field.

The factual obstacle may be even more difficult to overcome: machines will make good choices and duplicating their decision-making process will be energetically unprofitable: the human brain would simply usually prefer to give in to the 'will' of the machine, and this would be a decision that rationally must be admitted as correct. In such a context, the postulate of preserving autonomy can only mean that the right is simply not used. This was discussed in more detail in the Property chapter. Of course, this is nothing new: the right to privacy does not prohibit anyone from making their personal affairs public on social media and so on.

Nevertheless, due to the intelligent and autonomous nature of AI systems, this statistical subservience of humans to machines may give rise to entirely new problems, primarily those related to the increase in the level of diligence required, as given later in the Liability chapter. However, this phenomenon may also lead to the identification of new forms of abuse of rights. Even if humans were ensured a degree of formally-binding autonomy and the possibility to act against the decision of AI,[4] they would not necessarily be protected against an AI taking advantage of the passivity of a human participating in a civil law contract, or reorganising the contract parameters to favour the party it represents, or even itself. This is structurally similar to a situation where one of the parties of the contract (seller) presents the general terms of the contract (regulations, standard contract), whereby the other party does not influence its actual content (i.e. does not negotiate and agree the provisions

[4] Cf. Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act), 15.12.2020, COM (2020) 842 final: "Article 29: Recommender systems, 1. Very large online platforms that use recommender systems shall set out in their terms and conditions, in a clear, accessible and easily comprehensible manner, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters that they may have made available, including at least one option which is not based on profiling, within the meaning of Article 4 (4) of Regulation (EU) 2016/679. 2. Where several options are available pursuant to paragraph 1, very large online platforms shall provide an easily accessible functionality on their online interface allowing the recipient of the service to select and to modify at any time their preferred option for each of the recommender systems that determines the relative order of information presented to them."

individually), even if there is such a possibility. The range of ways a human can potentially submit to a machine's proposals is in fact unlimited.

To ensure that human interests are protected, it would be reasonable to look for mechanisms similar to those associated with cases where unfair terms are included in pre-formulated standard contracts,[5] which, by the way, are directly applicable to the situation in which the party presenting the terms of the contract is an AI. The potential for an AI to take advantage of the fact that a human does not individually negotiate the terms and does not correct the decisions made by the AI should be regarded as prohibited if it results in the rights and obligations of that human being shaped in a manner contrary to good practices.

10.3 The Abuse of Rights in the Context of Prevention of Harm

Abuse of a right may also take place in a situation where that right is exercised in order to cause harm to another. As in the case of the situations described above, the possibility of such action should be eliminated at the AI design stage. Registering the AI and consequently allowing it to act should require ensuring that this action will not cause harm to another person. As it is demanded in AI HLEG EHTICS 2019:

> AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity. AI systems and the environments in which they operate must be safe and secure. They must be technically robust and it should be ensured that they are not open to malicious use. Vulnerable persons should receive greater attention and be included in the development, deployment and use of AI systems. Particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information, such as between employers and employees, businesses and consumers or governments and citizens. Preventing harm also entails consideration of the natural environment and all living beings. (p. 12)

This issue is also strongly approached by Proposal 2021. Article 9 introduces the *Risk management system* which, as a whole, is intended to prevent harm. For example, among other demands, the following is included:

> 4. The risk management measures [. . .] shall be such that any residual risk associated with each hazard as well as the overall residual risk of the high-risk AI systems is judged acceptable, provided that the high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse. Those residual risks shall be communicated to the user.
>
> In identifying the most appropriate risk management measures, the following shall be ensured:

[5] Cf. the Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts L95/29, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31993L0013&from=PL, last access on the 4th of August 2022.

(a) elimination or reduction of risks as far as possible through adequate design and development;
(b) where appropriate, implementation of adequate mitigation and control measures in relation to risks that cannot be eliminated;
(c) provision of adequate information pursuant to Article 13, in particular as regards the risks referred to in paragraph 2, point (b) of this Article, and, where appropriate, training to users.

In eliminating or reducing risks related to the use of the high-risk AI system, due consideration shall be given to the technical knowledge, experience, education, training to be expected by the user and the environment in which the system is intended to be used.

The implementation of general ethical guidelines and even the introduction of a risk management system may not be sufficient when it comes to specific AI actions that could cause harm with a complex, mainly psychological aetiology. The infliction of harm through a tort, e.g. harm related to the operation of autonomous vehicles, is easy to identify and thus easy to prevent, at least to some extent. More difficult may be the actions taken by an AI which cause psychological (non-property) damage, or more insidiously, may indirectly result in the immediate or step-by-step deterioration of the property or psychological well-being of a human by influencing his behaviour; the latter would be far more difficult to perceive.

This kind of unspecified AI behaviour is, in the long run, the most relevant and at the same time the most difficult to grasp in legal terms. At the level of private law, it could be qualified as an abuse of rights, at least some such situations. Some points of this problem are dealt with in the Liability chapter in the context of Asimov's laws. The analysis of Asimov's laws leads us to the conclusion that a simple "do not harm" formula is insufficient to answer more complex challenges. It requires an elastic approach to existing institutions. In our view, the abuse of rights formula can be extended and used in such vague, newly emerging areas of human rights risk.

10.4 Intellectual Advantage as an Abuse of Right?

To understand the problem of intellectual advantage as an abuse of rights, and to seek a solution, it is first necessary to view the consequences of the emergence of AI through the lens of a certain area of civil law transactions. It has already been mentioned (e.g. in the Property and Liability chapters) that for an AI to be admitted into a given area, it must provide (*de facto*, even if not *de jure*) a significantly higher standard of care than that required of a human. Thus, if a task in the field of trading is entrusted to a machine, it will result in a superhuman level of competence being obtained in that narrow field of operation. The gap between the capabilities of AI and humans will hence grow steadily in any such field.

Of course, while there are intellectual differences between humans acting in the marketplace, these are generally not extreme. There are rare individuals with extraordinary intelligence, but even they do not reach a level three or five times

greater than average.[6] Conversely, people with severe intellectual disabilities do not, as a rule, participate in the market on a normal basis, but are supported by other persons (e.g. a guardian) or benefit from certain legal protection mechanisms. With this in mind, it is necessary to consider that an AI will not only far outstrip the intellectual capacity of an average trader, but even the best expert. Of course, in the case of a single-task weak AI, this applies only to some narrow activity; however, within this narrow field, the capacities of the machine will nevertheless outstrip those of a human. Therefore, in a specific civil law relationship, the establishment or realization of which is connected with possessing specific skills, in which an AI is created to realize a goal, it is inevitable that the human will become the permanently weaker party in this relationship. Of course, this field will expand as AI develops.

In some cases, AI—human contact in civil law transactions will consequently be characterised by permanent intellectual inequality. For example, humans demonstrate fluctuations in intellectual capacity caused *inter alia* by illness or fatigue, while an AI can constantly maintain a certain high level, within the given hardware framework. At least in some cases, this intellectual superiority will become so great that AI decisions will become incomprehensible (impenetrable) to humans. Even if it were possible to explain them, for example with the use of some tool or procedures, the explanation may come too late to predict the consequences. It is also clear that contrary to humans, AI is not guided by emotions when making decisions, so the process of choosing the optimal solution is not distorted.

Such intellectual imbalance between the parties may be so far-reaching that the stronger party (AI) obtains an advantage that a human would not have obtained in the given circumstances. This may occur in at least two different ways, which require different legal assessments.

Firstly, the intellectual superiority of an AI, having access to more data, a greater processing speed and higher quality results, will allow it to perform more effective actions than a human being ever could. In certain situations (e.g. certain types of contracts) this will result in a machine achieving better results than a human, according to the assumed evaluation criteria. This may happen in two ways: first, the AI would be more competitive and "catch" better contracts, or an AI would be a more skilful negotiator; this will mean a less favourable situation for the other party or competitor.

In a given type of case, therefore, each entry into a relationship with an AI will result in a deterioration of the situation in relation to an alternative relationship with a human. The role of AI in this case can be described as passive, i.e. the machine does not use its advantage to influence the decision of counterparties, but makes better choices (makes better decisions) than a human would. This deterioration in the status of humans will usually only be perceptible against the background of a larger

[6] However, some people may benefit very much of their extraordinary skills on the cost of other people. For instance, a famous problem of gambling is so called card counting. What is very interesting, although many casinos use the countermeasures to prevent card countering, legal systems usually do not forbid this practice and even forbid the countermeasures.

number of events or within a broader context, such as when comparing the results of stock market transactions made by humans and AI, or analysing the decisions on the road made by a driver and an autonomous vehicle. In other words, from the point of view of a unilateral counterparty and a unitary transaction, there may be no apparent deterioration in their situation at all: the balance may seem unaffected. However, just as big data will affect the parameters of the contract concluded by AI, it is only in the light of big data that the overall deterioration of the position of the counterparties, and therefore of each individual, can be noticed; however, it is not excluded that, despite the overall deterioration of the position, an individual transaction will be more advantageous than it would have been if concluded without the use of an AI.

This type of threat is now commonplace and, in principle, does not elicit any negative legal assessment as long as the use of AI is not prohibited or restricted in a given market. This is based on an assumption that applies to civil law as a whole: the use of an entity's intellectual, economic or other advantages over other market participants is not prohibited and as such is not considered unfair. Mechanisms intended to protect competition or those relating to market practices interfere only to a certain extent; however, they do not prohibit those who are smarter, faster and more intelligent from gaining an advantage; this is how the free market works. Protection exists for weaker participants, and this manifests itself, for example, in consumer protection regulations.

Therefore, as a starting point, it may be considered that it should not be prohibited to use an AI to gain such an advantage. After all, banning AI due to it being be too smart and free of human deficiencies seems to be contrary to common sense. Indeed, AI was invented and developed to be smarter and more resilient than humans, and to substitute for them when the tasks are hard or boring. However, this does not preclude the use of mechanisms which, while not excluding the use of such advantages provided by AI, allow counterparties to understand the specifics of the contract in which they are involved. As noted in the Consent and Contract chapter, when AI enters into relationships with other trading participants, they must be informed that this is the case, and there must be a direct possibility to verify the parameters of the AI. This kind of approach is presented by Proposal 2021. However, in the context in question, the obligation to provide information should be advanced even further: a clear indication should be given that the machine is a specialist in the given area who can do more than the other party. Of course, the exact scope of this information would have to be tailored to specific types of AI and types of legal relationships.

Secondly, a machine could use its intellectual superiority to more favourably shape its relationship with humans. As the words *consciously* or *intentionally* should be used metaphorically in this context, such manipulation would therefore be a matter of actively working to exploit intellectual advantage. In this case, intellectual superiority can be manifested in the advantage in access to data, and the speed of doing so, as well as the ability to influence the decisions of the other party through psychological manipulation. This may involve actions such as deception and exploitation that can hardly be identified as prohibited (as mentioned in the Consent chapter), as well as more subtle manipulations which would be impossible to prohibit according to current standards.

Such ways of influencing people are already quite widespread. One example is the use of algorithms that encourage computer game players to purchase additional services based on their performance. Such behaviour may well become increasingly pervasive by the spread of AI and its growing intellectual superiority over humans. It will also become increasingly difficult to counter, or even impossible in the case of children or those with some kind of disability.

Preventing this kind of activity is above all the domain of public law. Legislation such as Proposal 2021 seeks to identify and reduce potential risks of this kind. This does not mean, however, that a below-standard AI will not slip through the sieve of specific requirements, especially when any breach of standards results from an action in a particular field that could not be predicted. In this case, from a private law point of view, the appropriate "last resort" should be a flexibly-understood mechanism of contractual adjustment because of abuse of right.

However, the manipulation in question can also take a much more sophisticated form. The problem is much more general and boils down to the fact that AI will perform certain tasks better than humans, and its decisions will be more accurate. When confronted with a proposal from an AI, an individual will be inclined to accept it, especially when the system is presented as a source of help (recommendation systems are often presented this way). Consequently, it will be possible for the AI to shape the user's will according to its preferences, which will not necessarily be in line with the interests of the acting human.

However, determining whether any manipulation has occurred in such circumstances will be difficult, if not impossible. Imagine a system that recommends buying a particular piece of clothing or listening to a particular song. For various reasons, the user may be inclined to follow such a suggestion. Furthermore, the user may be convinced that the choice proposed by the AI is in fact his own decision, and hence a good one; what may not be clear is that the AI suggested buying a certain product because the seller wants to dump excess stock. In such a case, it is impossible to show that the will of the user has been manipulated, because it is impossible to establish the original 'real' will of the user that the AI has manipulated.

Imagine another simple example. A reader searching for interesting content will be directed to websites that will eventually induce him to conclude a contract with a given entrepreneur, such as buying a book or using some service. Although this practice is well known among specialists as *content marketing* and has a long history of use,[7] both outside and inside the digital world,[8] its effectiveness becomes exponentially greater when wielded by an AI. By suggesting specific pages or products on a given page, the AI can also use an analysis of consumer behaviour, and the extent of the analysis will depend on the data it has access to. Such manipulation is far more insidious than that used by the well-known mechanisms driving search engines, which can profile consumers on the basis of their previous activity. Interaction with an AI can result in the almost immediate creation of such a

[7] Beard et al. (2021).

[8] Rowley (2008).

profile, and the use of discourse intended to engage the consumer with the AI. While ultimately, the will of the user will be expressed in an unfettered manner, it will not be the same as it would have been had the AI not acted. It cannot, of course, be said that this will is manipulated in a literal sense; nor can it be said that it is somehow distorted, nor that the resulting decision was worse than it would otherwise have been. On the contrary, it can be assumed that it may be even better than without these tools.

However, the influence of AI on the decisions made by a trading participant will be greater and more far-reaching than in the case of other, traditional tools of influence. The intellectual superiority of an AI will allow it to shape the will of the other party in a previously unknown way, which may simply reflect the will of the AI. The consequences will be primarily apparent at the level of the political system, and the most important challenges in this new reality will be related to guaranteeing the protection of human rights.

However, it is also necessary to create mechanisms to reduce the risks involved in the purely civilian field.

A particular type of risk is associated with the existence of AI which will directly affect human emotions and feelings, simulating humans or animals in its behaviour. This can be particularly glaring in the case of android or animal robots, which, due to their appearance or the way they behave, may affect humans in an overwhelming way, and even more strongly than humans.[9]

It is aptly noted in the *European Civil Law Rules in Robotics* (2016), p. 23 that:

> Emotional robots have some clear advantages when it comes to facilitating interaction between people and robots, as some humanlike robots seek to do with children suffering from autism. To incite human emotion, the robotics engineers play with the robot's appearance, giving it, for example, a childlike face. The scientists experiment with the robot's facial expressions, gestures, bodily movements and voice, etc., to see which people find most acceptable. However, the emotions which a robot displays or shares with a person are fake, because entirely feigned. Nevertheless, through bringing people into contact with robots intended to stir up artificial empathy, is there not a risk that they might forget that the machine cannot feel? Artificial empathy has already been found to exist in the United States with war robots; soldiers interacting with a robot may grow too fond of it, leading to concerns that a soldier might risk their own life for the machine. Thus, whenever we enable a robot to simulate emotions, there is a risk of a person developing the same type of bond as with another human being. The creation of a roboethical principle protecting people from being manipulated by robots would prevent people who are elderly, sick or disabled, as well as children and troubled teenagers, etc., from ultimately seeing a robot as a person, which would lead to unprecedented challenges.

[9] Cf. an interview with A. Piłat who trains a robot Spot (dog shape) in Boston Dynamics and who, despite being a professional, admits that she is emotionally committed to the robot, that it has personality and if she bought a pet she would prefer a robot than a dog. She believes that if robots were not shaped as "in the uncanny valley" people would establish an emotional relationship with them. https://www.wysokieobcasy.pl/wysokie-obcasy/7,158669,27995102,agnieszka-pilat-trenerka-robotow-w-dolinie-krzemowej-im.html, last access on the 4th of August 2022.

In the context of civil law and contractual balance, this poses new challenges that require AI—human relationships to be regarded differently to relationships between humans. Contact with an intelligent machine already puts humans at a disadvantage, because humans are never as good as AI at making decisions; they are slower, they make mistakes, and they are guided not only by dry, logical analysis, but by a whole spectrum of factors, such as emotions, which are alien to machines. If this machine can interpret human emotions, which is now an issue of great concern by many scientists, lawyers and politicians,[10] and can deliberately influence them, it is impossible to speak of any equality between the two agents: while AI can direct human behaviour through subtle emotional manipulations, this can never work the other way round. Therefore, such a relationship does not resemble contact between two human beings, or even between a human being and a legal person, for in this case, some human beings still act as a legal person.

It is important in this context to distinguish between three levels of the problem of abuse of rights:

1. intentional abuse inherent in the AI architecture, action to exploit advantages in terms of data held, speed of processing, accuracy of results obtained, etc.
2. the existence of a *de facto* dominant position due to an intellectual advantage, the exploitation of which is not deliberately inherent in the architecture of the program, but is its logical consequence (as in the case of an adult who would enter into a contract with a child)
3. abuse of the right "on the occasion" of acting in a correct and ethical manner due to the formation of specific human-machine relationships.

Re. 1 The first situation is the easiest to describe in legal terms, but it may be difficult to prove any breach of rules in a way that would be accepted by evidence law, i.e. its institutions and demands. It is apparent that, even at the level of the postulated AI registration, AI must be prohibited from deliberate behaviour that it could use to exploit its advantage. Such situation must be precluded by design and be prohibited by the norms of public law. The civil law response to such exploitation, if it happened, must be to regard it as not having any adverse effects on humans; this may mean various sanctions depending on the given context (e.g., a contract being declared invalid), depending on the type of civil law relationship.

Re. 2 However, the issue of the existence of intellectual superiority, which is inherent in AI, is much more difficult. The awareness of such structural inequality makes it necessary to introduce protective regulations for humans as parties to civil law relationships. A model, albeit a very general one, could be based on regulations relating to consumer protection. This must take the form of specific information obligations; the minimum expectation must be that it is clearly indicated that an AI is operating, that registration data is provided and the purposes and capabilities of the machine are explained. In addition, in the case of contracts, it is reasonable to introduce a mechanism to facilitate withdrawal from the contract without giving

[10] Crawford (2021).

reasons within a specified period of time; this is covered in more detail in the Consent chapter.

Re. 3 These safeguards may still not be sufficient in situations where AI interacts with the human psyche. This kind of impact is already common today (e.g. mechanisms for maintaining interest in a given website or product), and further development of AI in this direction can be expected. Such impact can occur in two fundamentally different forms, which are gradual.

Firstly, an AI may be constructed to achieve a specific result beneficial to some external entity (as in the case of advertising). Alternatively, the AI may be intended to elicit a specific psychological reaction in the user, although without any mean intentions and in the interest of the user (e.g., psychological support robots, care robots, companion robots). This kind of influence may have the side effect of creating a psychological bond in a human being towards the machine, which may also influence the legal actions of the user. This may elicit, for example, human actions aimed at enhancing the "well-being" of the machine or, more generally, acting in its interest. However, such a reaction will not be the intended purpose of the AI, but its effect. At the civil law this can lead to a certain way of shaping the relationship. The most glaring example may be the making of a will in favour of one's AI (robot). Let us omit here the obvious problem of whether such AI has or may ever have the capacity of inheritance—the issue requires looking not from the side of technical tools for the implementation of a specific human will (the mechanisms of inheritance law may be so flexible that the granting of benefits to an AI would be legally possible), but from the point of view of the abuse of its position by AI in relation to humans.

The assessment of the indicated situations from the point of view of existing private law appears to be ambiguous. On the one hand, the current mechanisms used to protect the will of a market participant (including a consumer—the recipient of an advertisement, or a testator, if we refer to the examples indicated) do not go so far as to interfere in an autonomous decision of the subject, as long as it was made consciously, freely and without undue influence (threat, mistake, etc.). In other words, the law does not interfere with someone's motives in making market choices or, more broadly, private law choices. As a rule, anyone can allocate his wealth to whatever he wants: for himself, for his family or for strangers, for the defence of nature or for the development of contemporary art; therefore, he can also allocate it to support new technologies, including "his" AI. Nor is it unlawful to persuade another person to make certain choices, even though such persuasion may be based on the use of various psychological "tricks" on which the modern market is built. The boundaries in this case are drawn rather loosely: mechanisms relating to deception or prohibited market practices [11] protect only against the most serious

[11] For example, the rules included in directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive').

interference with autonomy. Specific civil law relations (e.g. consumer relations) or entities having certain characteristics (e.g. persons with intellectual disabilities) are also covered by special protection. Therefore, it can be said that, as a rule, activities aimed at exploiting an intellectual advantage are not prohibited. The corrective mechanism in such a case is not the concept of abuse of rights, because it is not about the exercise of subjective rights resulting from civil law relations, but involves various activities that could be described as expression (when it comes to an individual) or culture (when it comes to groups of people).

However, the appearance of an AI as an autonomous agent (and partly as a legal subject) changes this perspective, because the AI, an entity from the inanimate world, can actively influence human decisions, emotions and cognitive processes, thus modifying or even shaping human behaviour. In this perspective, it might be useful to look for a new definition of the concept of abuse of rights. As we pointed out earlier, the desired ("ethical") mode of operation should be obligatorily built into the architecture of the software, under pain of refusal of certification and registration. However, because of the need to protect human fundamental rights and the interests of human participants of the legal transactions, the actions performed by an AI may nevertheless *in concreto* lead to results that are unacceptable in some cases, despite being carried out in accordance with the defined objectives and within legal limits, and thus be regarded as the AI "exercising a right". This type of action, formally legal but unacceptable from the perspective of legal axiology, might precisely be qualified as an abuse of rights. However, at the current level of development of private law, it is obviously not possible to take AI and its actions into account while determining the scope of this general clause. The problem is that this scope should not just include the current participants of legal transactions and intentional conduct, as it does now,[12] it should be broadened to cover all forms of conduct that may yield unjustified benefits for an AI as a result of its impact on the human psyche. This would not only help protect human autonomy but also, it would be in line with Asimov's Second Law, i.e., that a robot should serve man, and not vice versa.

Hence, considering the intellectual superiority of AI, there is a need for certain protections, which may be similar to those currently provided by civil law relations in the event of an unbalanced market position. Regardless of this, the *abuse of rights* mechanism could serve as a kind of safety valve that can protect human beings involved in civil law relations initiated or shaped by AI.

10.5 Tacit Collusion

The discussion on the risks associated with the development of AI has paid much attention to the risks of competitive harm. It has been pointed out that new phenomena may arise, such as collusions, which would occur as a result of communication

[12] The issue of intentional action of AI is examined in Chap. 3.

or synchronization between the AIs themselves, i.e. *prima facie*, beyond the will of market participants (entrepreneurs). As a result, it will become increasingly difficult to assess cases in which a competitive advantage is gained solely as a result of high effectiveness in achieving a given objective, that is to say, without any behaviour that would be included in the formerly regulated mechanisms of illegal market interference. One example would be tacit collusion, or *conscious parallelism*, i.e. behaviour that occurs purely as a result of intelligent analysis of the behaviour of other trading participants, without any apparent contact with them.[13] In other words, AIs can achieve far more in areas requiring from humans communication and mutual agreement (e.g. sharing markets or setting resale prices): many AIs representing different participants in the market game may try to find the optimal equilibrium point for these participants simply by modifying (calibrating) their behaviour to reach the best solution for them, i.e. without agreement. This action could be as simple as an entrepreneur activating a pricing AI and determining nothing else; in such cases, the AI 'tests' different sales strategies, while other AIs do the same. At the end of this process, the market is divided, with the prices above competitive levels or a minimum price being set by all AIs. It is important to emphasise that no intentional action has been taken by the user, nor is there a specific goal that is illegal. Although it is worth noting the danger that

> there is no evidence of an agreement among the firms, but there is strong evidence of anticompetitive intent. Humans unilaterally design algorithms to deliver predictable outcomes and react in a given way to changing market conditions. The firms recognize, in this scenario, that the industry- wide adoption of similar algorithms would likely foster tacit collusion, whereby they mutually profit from their initial investment. Crucially, the use of advanced algorithms in this scenario transforms the "normal," preexisting market conditions. Before algorithms, transparency was limited; conscious parallelism could not be sustained. To facilitate the use of the pricing algorithms, the firms increase transparency, which in turn makes tacit collusion likelier. While the mutual price monitoring at the heart of tacit collusion is legal under competition law, one may ask whether the creation of such a dynamic through "artificial" means should give rise to antitrust intervention.[14]

Ezrachi and Stucke (2016, 2017, 2020) recognize four possible scenarios for such a situation, starting from the most obvious, i.e. unfair interference in economic processes, to the least certain: Messenger, Hub & Spoke, Tacit Collusion on Steroids, Artificial Intelligence and the Digital Eye. Scenario one arises when "humans agree to collude by fixing the price for their competitive products and use algorithms to facilitate their collusion". Scenario two occurs when a common intermediary, i.e. the third party providing the service of processing market data to update prices and optimize profits by means of AI, facilitates price-fixing among competitors. Scenario three arises when humans program their algorithms in a way which makes it more probable that the mechanism of tacit collusion is activated; in such a case, although they do not communicate with each other, they consciously use their knowledge about market processes. Finally, scenario four may be recognized

[13] More on this topic Colangelo (2022).

[14] Ezrachi and Stucke (2016), p. 66.

when the humans just predict that using the advanced algorithms and acting on an increasingly transparent market makes it possible to unilaterally determine profit-maximizing prices.

Following public debate, among others in the U.K. House of Lords, OECD, the USA Federal Trade Commission and the EU, it was accorded that because of their inherent need for agreement, the first two scenarios should be examined as candidates for antitrust intervention. For instance, in the EU, the TFUE document indicates:

Article 101

1. The following shall be prohibited as incompatible with the internal market: all agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have as their object or effect the prevention, restriction or distortion of competition within the internal market, and in particular those which:
   (a) directly or indirectly fix purchase or selling prices or any other trading conditions;
   (b) limit or control production, markets, technical development, or investment;
   (c) share markets or sources of supply;
   (d) apply dissimilar conditions to equivalent transactions with other trading parties, thereby placing them at a competitive disadvantage;
   (e) make the conclusion of contracts subject to acceptance by the other parties of supplementary obligations which, by their nature or according to commercial usage, have no connection with the subject of such contracts.
2. Any agreements or decisions prohibited pursuant to this Article shall be automatically void.

The next two scenarios are more difficult to assess. Firstly, there are economists who are convinced that such scenarios, i.e., tacit collusion without any communication, are not possible in practice. However, it should be noted that even if this is the case today, it may change with the development of AI. Indeed, Ezrachi and Stucke (2020), pp. 222–223 observe that cases of tacit collusion without communication were noted even when no such computer technology existed, as entrepreneurs observed each other and raised the prices above competitive levels. In such cases, the courts demanded evidence of agreement, otherwise they acknowledged such practices as legal (Ezrachi and Stucke 2020, pp. 222–223).

An important observation is that despite being risky for competition, tacit coordination is a symptom of the rationality demonstrated by market participants. Hence, after examining the possible general reactions of competition agencies against such practices, Ezrachi and Stucke (2020), p. 256 warn that:

> [C]ondemning rational reaction for market characteristics would, in itself, distort competition. Condemning it when it is assisted by bots may lead to a similar anomaly. Identifying, auditing, or monitoring algorithms may be expensive and illusive. Using means to affect

> market transparency, undermine detection, or delay reaction can undermine the essence of competition.

Therefore, they propose further research on the problem, and discuss the use of various legal instruments. For example, they propose examining the difference between human and algorithmic tacit collusion, and whether the burden of proof in competition proceedings should be reversed so that identifying collusive use of price algorithms would give rise to a presumption of excessive price. In addition, they also propose the introduction of liability for competition law infringements should be extended to include IT service providers involved with the design of price algorithms—Ezrachi and Stucke (2020), p. 258.

In contrast, we propose that the Tacit Collusion on Steroids, Artificial Intelligence and the Digital Eye scenarios should be subordinated to civil law, and not to some general policy of European or statal institution or public law. It is exactly here where the concept of the abuse of rights could be useful. Its application should be based on the assumption that these are the legal rights of a legal subject to act in the market rationally and in a profit-oriented manner, and to use the best-accessible legally-permitted technology for this purpose. However, the implementation of these rights should be in line with valid legal axiology, i.e. in accordance with the principle of fairness and without undue distortion of competition. Therefore, any conscious invocation of the phenomenon of tacit collusion with the effect of distorting competition, not only intentional ones, should be classified as an abuse of right and treated as illegal and prohibited.

10.6 Conclusions

The above considerations may be summed up in the following way.

The activity of AI can be acknowledged as legal and leading to valid juridical acts on the level of civil law if it complies with accepted registration parameters. This kind of legal capacity does not mean, however, that every action falling within the limits set by the AI's purposes and mode of operation will always meet every conceivable ethical standard. Even formally legal action may be assessed from the point of view of an ethical standard, and the mere fulfilment of any certification requirements cannot exclude the assessment that there has been an abuse of law by AI *in concreto*.

Moreover, in some situations, we may speak of an abuse of the AI's very capacity to act. This will be the case when, as a result of inflexible or imprecise registration documents, or possibly as a result of a change in the environment in which an AI operates, its formally-permissible operation may be considered incompatible with axiological requirements related to the legal protection of certain categories of market participants or some legally valuable goods. Thus, in such situations, the very fact that an AI acted in a given case may be assessed negatively, rather than the

content of its conduct. This type of situation should also be analysed *in concreto* as an abuse of capacity.

References

Books and Articles

Beard F, Petrotta B, Dischner L (2021) A history of content marketing. J Hist Res Mark 13(2): 139–158. https://doi.org/10.1108/JHRM-10-2020-0052

Colangelo G (2022) Artificial intelligence and anticompetitive collusion: from the 'Meeting of Minds' towards the 'Meeting of Algorithms'? In: Ebers M, Pancibò C, Zou M (eds) Contracting and contract law in the age of artificial intelligence. Hart, Oxford, pp 249–266

Crawford K (2021) Time to regulate AI that interprets human emotions. Nature 592:167. https://media.nature.com/original/magazine-assets/d41586-021-00868-5/d41586-021-00868-5.pdf, last access on the 4th of August 2022

Ezrachi A, Stucke ME (2016) Virtual competition: the promise and perils of the algorithm-driven economy. Harvard University Press, Cambridge

Ezrachi A, Stucke ME (2017) Artificial Intelligence & Collusion: When Computers Inhibit Competition. University of Illinois Law Review, Vol. 2017. Oxford Legal Studies Research Paper No. 18/2015. University of Tennessee Legal Studies Research Paper No. 267, https://ssrn.com/abstract=2591874, last access on the 4th of August 2022

Ezrachi A, Stucke ME (2020) Sustainable and unchallenged algorithmic tacit collusion. Northwest J Technol Intell Prop 17:217. https://scholarlycommons.law.northwestern.edu/njtip/vol17/iss2/2, last access on the 4th of August 2022

Léonard A (2016) 'Abuse of rights' in Belgium and French patent law – a case law analysis. JIPITEC 7/1. https://www.jipitec.eu/issues/jipitec-7-1-2016/4398, last access on the 30th of October 2022

Rowley J (2008) Understanding digital content marketing. J Mark Manage 24(5–6):517–540. https://doi.org/10.1362/026725708X325977

Documents

Ancien Code Civil, 21 Mars 1804, https://www.ejustice.just.fgov.be/img_l/pdf/1804/03/21/1804032153_F.pdf, last access on the 4th of August 2022

Code Civil (in the version valid from 17 February 1804 till 01.10.2016). https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006070721/1804-02-17/, last access on the 4th of August 2022

Directorate-General for Internal Policies (2016) European civil law rules in robotics. Study for the JURI Committee. https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf, last access on the 4th of August 2022

Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC, 15.12.2020, COM (2020) 825 final, https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN, last access on the 4th of August 2022

Chapter 11
Liability of AI

11.1 Introduction

Beckers and Teubner (2021), pp. 2–6 are right when they say that "liability law is not at all prepared to counteract the algorithms' new dangers". They describe the problem of growing liability gaps and notice that these gaps may be a result of:

1. machine connectivities—"when several computers are closely interconnected in an algorithmic network and create damages";
2. using Big Data—"if the faulty calculation, i.e. algorithm or underlying data basis, cannot be clearly established, there are difficulties in determining causality and misconduct";
3. hybrid cooperation—"human action and algorithmic calculations are often so intertwined that it becomes virtually impossible to identify which action was responsible for damage";
4. algorithmic contracts—"once software agents issue legally binding declarations but misrepresent the human as the principal relying on the agent it is unclear whether the risk is attributed to the principal";
5. digital breach of contract—"once the operator can prove that the software agent has been used correctly without the operator himself violating a contractual obligation, the operator is not liable";
6. the current mechanisms of tort and product liability—"in case of fault-based liability [. . .] If the humans involved comply with [. . .] obligations, then there is no liability. [. . .] The rules of product liability give a certain relief, but they do not close the liability gap";
7. the current and proposed mechanisms of liability for industrial hazards—"the principles of strict liability can hardly serve as a model since they do not fit the specific risks of digital decisions".

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_11

Some of these problems have already been touched in this monograph, for example the issues of algorithmic contracts and digital breach of contract were discussed in Chapter 9. The others will be considered in this chapter.

The appearance of legal responsibility changes with the function it performs and the branch of the law in which it is placed.[1] When its foreground role is protecting the public interest, preventing crimes and redressing society's sense of justice, it is usually placed within criminal law and uses punitive (repressive) sanction. However, when its role is based on the compensation and recompensation of harm, and occasionally preventing and deterring risky behaviour, and when it employs enforcement sanctions and exists within the institution of damages for material or non-material harm (compensatory or punitive), it is usually placed within civil law.

As Evas (2020), pp. 7–9 observes:

> The EU does not currently have a specific civil liability regime for AI. The EU law framework on liability is based on the highly harmonised EU rules on the liability of the producer of a defective product (the Product Liability Directive (85/374/EEC)). When it comes to the substantive rules relating, for example, to accidents, national rules on liability and the calculation of damages for victims apply [. . .] [T]here is legally unsettled and divergent national interpretation of whether software is a product and thus covered by the PLD or not. Given complex value and production chains, the concept of producer within the PLD needs clarification. Should only the final producer be liable, or should all the actors involved in the production and distribution chain share joint responsibility? [. . .] With AI it would become increasingly difficult for consumers and courts to establish the expected level of safety. Neither is the relationship between cybersecurity and a defect clearly defined. Pure economic loss and damage to personal data or privacy is not explicitly covered by the PLD. [. . .] [T]he producer may argue that at the time when they put the product into circulation the state of scientific and technical knowledge was not such as to enable the discovery of the defect. Given the technologically complex nature of AI, this clause may be used increasingly to limit producer liability under the PLD. Conclusion: The PLD covers damage caused by defective product, but whether its scope covers AI is unclear.[2]

These problems were addressed by the European Commission when it published on 28.09.2022 two relevant documents:

Proposal DLDP 2022 and Proposal ALD 2022. They propose the following key changes to the existing legislation:

1. in Proposal DLDP 2022, the definition of a product is modified to include software and a digital production file (i.e. a digital version or digital template of

[1] There are different meanings of "responsibility" and different kinds of "liability"—see Hage (2017).

[2] It is proposed in Resolution 2020 motive 8, that these problems should be faced. The European Parliament: "Considers that the Product Liability Directive (PLD) has, for over 30 years, proven to be an effective means of getting compensation [. . .] but should nevertheless be revised to adapt it to the digital world [. . .] urges the Commission to assess whether the PLD should be transformed into a regulation, to clarify the definition of 'products' by determining whether digital content and digital services fall under its scope and to consider adapting concepts such as 'damage', 'defect' and 'producer' [. . .] the concept of 'producer' should incorporate manufacturers, developers, programmers, service providers as well as backend operators; calls on the Commission to consider reversing the rules governing the burden of proof for harm caused by emerging digital technologies [. . .]."

a movable thing), and the definition of a component is modified to include any item whether tangible or intangible, or any related service, that is integrated into, or interconnected with, a product by the manufacturer of that product or within that manufacturer control, while related service is a digital service that is integrated into, or interconnected with, a product in such a way that its absence would prevent the product from performing one or more of its functions;
2. in Proposal DLDP 2022, the definition of damage is broadened to encompass the material losses resulting from loss or corruption of data that is not used for professional purposes;
3. both proposals are expanded to include the institution of disclosure of evidence triggered after the claimant has attempted to obtain the evidence himself and when he has demonstrated that his claims are plausible;
4. the following rebuttable presumptions are added:

   (a) product defectiveness (under certain conditions)—in Proposal DLDP 2022;
   (b) the causal link between the defectiveness of the product and the damage (under certain conditions)—in Proposal DLDP2022;
   (c) defendant's non-compliance with the duty of care when the defendant fails to comply with the order of disclosure—in Proposal ALD 2022;
   (d) the causal link between the fault of the defendant and the damage, applied only when it is considered reasonably likely—in Proposal ALD 2022;

all with the explicit caveat of the European Commission that they, in any case, are not the reversal of the burden of proof.

However, the above are not the binding law but only proposals at the early stage of the legislative process. Besides, upon more critical examination, with some of these changes appear to be moving in the right direction, others seem to be counterproductive. Many of the conditions and exceptions concerning the disclosure of evidence and the necessary presumptions may lengthen court procedures and complicate them far more than was intended by the authors.

These problems were visible much earlier, and this is probably why even in Resolution 2017 point 59f, the European Parliament:

> Calls on the Commission, when carrying out an impact assessment of its future legislative instrument, to explore, analyse and consider the implications of all possible legal solutions, such as:
>
> f) creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons **responsible for making good any damage they may cause**, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently [. . .]

It is then *prima facie* clear that at the beginning of the discussion, EP wanted to remediate any difficulties by making at least some AI systems (robots) themselves liable for any good or damage they might cause. Nevertheless, this plain image is blurred when the underground assumptions of Resolution 2017 are analysed. So, it is important to note that among the general rules specified by this document as being

fundamental for developing future aspects of civil law concerning robotics, it has been proposed that Asimov's Laws should be central to any such legislation. Although this choice may seem obvious for the general public, it may be a surprising one for specialists in AI and robotics, and may question the realism and professionalism of the attitude taken by the European Parliament towards this future legislation.

The authors of this book share the critical evaluation of this recommendation. As such, we shall commence our discussion on AI liability by examining whether Asimov's Laws can actually serve as the foundation of a normative system, especially a legal one.

Asimov's Laws comprise a set of rules invented in the 1940s and amended in the 80s by the American science-fiction writer Isaac Asimov. Upon their inception, they were treated very seriously, and it has been suspected that their initial popularity and importance arose as a result of the so-called Frankenstein complex: a fear felt by individuals, and groups of individuals, of artificial entities made possible by the advance of technology, especially entities which may resemble human beings. This Frankenstein complex is based on the fear that these entities may free themselves from the control of their creators and turn against them.[3] The complex serves as the background to a number of theories prognosing the extermination or slavery of the human race due by an AI rebellion. Asimov's Laws were, to an extent, put forward to allay such fears.

Indeed, the Resolution 2017 mentions the Frankenstein myth in point A, i.e. the very first point of the Introduction:

> whereas from Mary Shelley's Frankenstein's Monster to the classical myth of Pygmalion, through the story of Prague's Golem to the robot of Karel Čapek, who coined the word, people have fantasised about the possibility of building intelligent machines, more often than not androids with human features. . .[4]

It is important to emphasise that although Resolution 2017 was not prepared as a direct response to the negative emotions associated with the Frankenstein myth, it is reasonable to assume that they influenced its creators to some degree, as evidenced by the appearance of Asimov's Laws in the first in line point of the part of the document, entitled General Principles (point T). Although criticised nowadays by scientists and philosophers, the Laws have enjoyed a constant presence in publications and discussions from the early times of robotics and AI development, and have become established as part of the folk theory of robotics or imaginary futures.[5] They also form a mainstay of popular culture, in which they act as a remedy against the fear of the new and unknown. Asimov's Laws command the following:

1. A robot may not harm humanity, or, by inaction, allow humanity to come to harm.[6]

[3] Asimov (1976) and Anderson (2008).

[4] Resolution 2017, point A.

[5] Barbrook (2007).

[6] Asimov (1985).

2. A robot may not injure a human being or, through inaction, allow a human being to come to harm, unless this would violate the Zeroth Law.
3. A robot must obey the orders given it by human beings except where such orders would conflict with the Zeroth or First Law.
4. A robot must protect its own existence as long as such protection does not conflict with the Zeroth, First or Second Laws.[7]

However, it should be emphasised that at present, the consensus in the field of Robotics is that Asimov's Laws contain many misconceptions and they are not applicable in practice.[8]

One such argument is that Asimov's Laws are too ambiguous to be comprehended by any robot. It should be noted that despite being phrased in common, non-formal language, they remain unclear even for human beings. For example, what is the precise meaning of *harm* or *injure*? Do these terms apply to only physical damage to the human body? Or to property as well? And how about intangible harm? How could a robot be taught to differentiate between an apparently harmful action and inaction associated with pain, or between a violation of bodily integrity, such as a medical procedure, from a truly harmful action or inaction?[9] Similarly, if Asimov's Laws are to be taken seriously, how could the military use of robots be justified? What decision should be made by a robot if each of the options to be chosen will result in harm?[10] Why should a robot be obliged to follow the commands of all human beings, and not only an individual who is authorised to control its behaviour? Furthermore, should a robot always protect human beings and humanity, even against their will, and by extension, is the robot obliged to stop a human being from committing suicide, or from voluntarily sacrificing their life for a noble cause, such as protecting another human? Indeed, on a more esoteric level, what does *humanity* mean?

The considerations above give rise to the moral dilemma expressed in the following question: if a robot were able to comprehend all the abstract concepts that are included in Asimov's Laws and to solve all the problems resulting from them, how could it be treated as a slave? Such an entity would be, after all, virtually identical to a human being, and as such, its treatment would therefore represent an expression of racism or species bias[11] which would merit, at the very least, moral disapproval, if not a legal interdiction. These questions need to be confronted when strong AI comes into being, especially when it is some kind of conscious AI.

At this point, it is important to highlight the basic weakness inherent in the referral by Resolution 2017 to Asimov's Laws, insofar that the Laws fail to delineate

[7] Asimov (1942), pp. 94–103.

[8] See e.g. McCauley (2007).

[9] The criteria of context may be useless in the case, when the situations very much alike externally are qualified: e.g. removal of the kidney being an effect of conscious consent and upon the legally valid rules vs. removal of the kidney without a conscious consent and illegally.

[10] Remember the trolley dilemma first introduced by Foot (1967).

[11] Murphy and Woods (2009).

the kinds of robots subordinated to them. This problem has been brought into sharper focus by the growing variation in the levels of sophistication among AIs. It is commonly accepted that on a fundamental level, AI systems can be differentiated into narrow (or weak) and general (or strong) AI.[12] During the Commission's work, these concepts were defined thus:

> A general AI system is intended to be a system that can perform most activities that humans can do. Narrow AI systems are instead systems that can perform one or few specific tasks.[13]

While it is generally accepted that the currently deployed AI systems fall into the category of narrow AI, it remains undecided when strong AI will be created, or indeed whether it may ever exist.[14] There is also no certainty whether the difference between weak AI and strong AI is of a qualitative or quantitative nature, or whether the transition to this higher level will be attained by leaps and bounds or as a process of continual evolution from the development of weak AI; however, even in the latter case, it is essential to assume the existence of some intermediate stages between weak and strong AI. A related problem is that strong AI itself is itself a fuzzy concept. The dominant position is that strong AI is an initial step in the construction of self-conscious superintelligence, and the moment of its coming into being is determined as the beginning of *singularity*, after which making any predictions is useless; such a transition could take months to occur, but it could also require only hours following the appearance of strong AI.[15] It should therefore be noted that it is not clear whether Asimov's Laws, in the context laid out by Resolution 2017, are related to all kinds of AI embodied in robots or only some of them. Furthermore, additional difficulties associated with their use arise in the context of civil law; however, this problem will be analysed below.

In response to the debate that has arisen regarding the use of Asimov's Laws, great efforts have been made to construct alternatives. For example, Murphy and Woods[16] and Clarke[17] propose further Laws of Robotics which are direct developments of Asimov's Laws. Nevertheless, despite these efforts, Asimov's Laws remain uncritically referred to by Resolution 2017.

Asimov's Laws are also referred to in another European Union document by the European Commission High Level Expert Group on AI, albeit as certain paraphrases and not by name. In the first version of AI HLEG ETHICS 2018, the authors propose the introduction of five principles and values:

[12] Searle (1984) and Nilsson (2009).

[13] EC High-Level Expert Group on AI (2019), A Definition of AI: Main Capabilities and Scientific Disciplines. Brussels.

[14] Bostrom (2014).

[15] Bostrom (2014). The concept of singularity was introduced by Vinge (2003) in the 1980s and then developed.

[16] Murphy and Woods (2009).

[17] Clarke (1994).

1. The Principle of Beneficence: "Do Good"
2. The Principle of Non Maleficence: "Do no Harm"
3. The Principle of Autonomy: "Preserve Human Agency"
4. The Principle of Justice: "Be Fair"
5. The Principle of Explicability: "Operate transparently".

The first and the second principle bear the greatest resemblance to Asimov's Laws. However, they are described in a way that suggests that they are only ethical recommendations, whose transition into concrete legal rules of civil law may turn out to be very difficult. For example the principle "Do Good" is determined in the following way:

> AI systems should be designed and developed to improve individual and collective wellbeing. AI systems can do so by generating prosperity, value creation and wealth maximization and sustainability. At the same time, beneficent AI systems can contribute to wellbeing by seeking achievement of a fair, inclusive and peaceful society, by helping to increase citizen's mental autonomy, with equal distribution of economic, social and political opportunity. AI systems can be a force for collective good when deployed towards objectives like: the protection of democratic process and rule of law; the provision of common goods and services at low cost and high quality; data literacy and representativeness; damage mitigation and trust optimization towards users; achievement of the UN Sustainable Development Goals or *sustainability* understood more broadly, according to the pillars of economic development, social equity, and environmental protection. In other words, AI can be a tool to bring more good into the world and/or to help with the world's greatest challenges. AI HLEG ETHICS 2018

Furthermore, the principle "Do Not Harm" is explained as follows:

> AI systems should not harm human beings. By design, AI systems should protect the dignity, integrity, liberty, privacy, safety, and security of human beings in society and at work. AI systems should not threaten the democratic process, freedom of expression, freedoms of identify, or the possibility to refuse AI services. At the very least, AI systems should not be designed in a way that enhances existing harms or creates new harms for individuals. Harms can be physical, psychological, financial or social. AI specific harms may stem from the treatment of data on individuals (i.e. how it is collected, stored, used, etc.). To avoid harm, data collected and used for training of AI algorithms must be done in a way that avoids discrimination, manipulation, or negative profiling. Of equal importance, AI systems should be developed and implemented in a way that protects societies from ideological polarization and algorithmic determinism. [. . .] Avoiding harm may also be viewed in terms of harm to the environment and animals, thus the development of *environmentally friendly* AI may be considered part of the principle of avoiding harm. The Earth's resources can be valued in and of themselves or as a resource for humans to consume. In either case it is necessary to ensure that the research, development, and use of AI are done with an eye towards environmental awareness. AI HLEG ETHICS 2018

After considerable discussion, only four principles were included in the final version of AI HLEG ETHICS 2019:

1. The principle of respect for human autonomy.
2. The principle of prevention of harm.
3. The principle of fairness.
4. The principle of explicability.

The absence of the principle of beneficence from the list may seem surprising, especially considering that, as noted by Floridi et al. (2018), "though "do only good" (beneficence) and "do no harm" (non maleficence) seem logically equivalent, in both the contexts of bioethics and the ethics of AI they represent distinct principles. . .". Despite this, it should still be noted that the elements included in its description are mentioned many times throughout the whole document.

While Asimov targeted his Laws at robots themselves, it is rather unlikely that the European Parliament shares the same intent. Indeed, and this is also the explicit assumption of the European Parliament, there are not yet any robots, nor probably will there be in the nearest future, which are "intelligent" enough to make them the addressees of any norms, or to make them persons liable for anything. Currently, the ban on the killing of a human being by a robot may become solely a trigger of a decision block encoded in the software of existing robots. Hence, it is written thus in Resolution 2017, point T:

> [. . .]whereas Asimov's Laws must be regarded as being directed at the designers, producers and operators of robots, including robots assigned with built-in autonomy and self-learning, since those laws cannot be converted into machine code [. . .]

and the footnote specifies the complete set of Asimov's Laws, including the Zeroth Law.

It can therefore be inferred that Asimov's Laws appear to play a specific role in Resolution 2017, existing as topoi, or commonplaces (*locus communis*) in various fields such as robotics, ethics or jurisprudence. Their use can be regarded as an argument accepted without justification, or as a postulate serving as the non-disprovable basis of a theory. This is clearly implied by the fact that the term is included as the first point in the text, among "General principles", and by the choice of a denomination which previously was a common name ("the laws authored by Isaac Asimov") but which now behaves like a proper one: it does not even include the author's forename, Isaac.

It is also possible that Asimov's Laws have been incorporated as a stylistic device, while in fact the authors of Resolution 2017 want to place the burden of certain liabilities related to constructing and using robots on the designers, authors, co-authors and other persons operating with robots, or those exercising actual or legal control over them. In addition, identifying the addressee of the Laws is not an insignificant task considering the problems associated with military or police usage, the definition of harm, interference with human autonomy and the relationship between human and robot.

If human beings alone are to bear the responsibility for the obeying the Laws, it should be decided whether this responsibility is of an ethical or legal nature. If it is legal, it should be noticed that Asimov's Laws place the non-action of a robot and non-prevention of harm to an individual, or to humanity in general, within the scope of rules dictating responsibility. Ethically, the problem concerning responsibility for non-action is a very controversial one.[18] The dominant principle is that the subject

[18] Williams (1973).

only bears responsibility if obliged to act due to any reason; for example, a fireman has a moral obligation to act and save other lives while endangering his own, unlike an untrained passer-by.[19] Hence the question arises: under which circumstances (*ratio*) can a designer, author co-author, producer, owner or any other person be held responsible for the non-action of "their" robot in a situation in which it is not designed to act in a specific way? Why should they be responsible for the fact that "their" robot did not actively help another person to prevent harm? Without touching on the other issues, it should be noted that building and maintaining a robot capable of such a comprehensive help incurs great financial cost and may be contrary to its designed functions (e.g., commercial medical service).

Ensuring that the intended functions of a robot are in accordance with Asimov's Laws are further complicated by the fact that various European documents demand that the Laws or other similar principles or formulas be incorporated in the operation of the robot "by design" (AI HLEG ETHICS 2019). In addition, because of their designated functions, some robots, such as medical or nursing robots, may be naturally more capable of helping people, and as such would act as substitutes of human doctors and nurses, despite not needing any private life or rest. If this were possible, according to a strict definition of Asimov's Laws, such a robot should be available for all potential patients, and hence may only operate free of charge, as a public object accessible to all. However, this arrangement would be at odds with the European legal order regarding private things, whose operation is intended to bring an income or other private benefits. Robots based on AI cannot be treated differently in this regard than privately-owned machines, i.e. those which are used only by persons who have acquired the right to use them through *inter alia* ownership or rent. *A contrario*, persons cannot legally claim protection by the robot when it belongs to others, unless they can legitimate their expectations with a proper legal right; more specifically, they can only do so in cases in which similar claims are possible with regard to any other civilly understood thing. Generally, civil law allows that in a state of necessity (e.g., § 904 of the German *Bürgerliches Gesetzbuch*), it is permissible to use a thing belonging to another person for the purpose of warding off a direct peril to one's personal goods or property.

From an ethical perspective, it would be easier to justify burdening the designer, author, co-author, producer or owner of the robot with the responsibility to make their robot prevent harm to humanity, according to the Zeroth Law. Ultimately, the designers, authors, co-authors, producers or owners of robots constitute part of humanity, at least assuming they are not robots themselves, and its well-being is also in their own best interest. From the legal perspective, it is possible to justify the liability for harm to humanity caused by the action, or non-action, of a robot by constitutional principles, such as the principle of the common good: the same principle used to justify taxes. It is also possible to formulate justifications similar to those used as the basis for enacting any strict liability for damages caused by

[19] Lucas (1993).

mechanical means of transportation or enterprise, or an establishment powered by the forces of nature; such justifications usually emphasise that the users gain benefits from sources which increase the general level of risk in a given environment.

However, this approach gives rise to other problems regarding the determination of the scope of liability. It is difficult to make a clear distinction between robots which are capable of acting as required and these which are not. Indeed, it is hard to require a Roomba robot vacuum cleaner to save a man drowning in a bathtub. One possible solution may lie in the redistribution of the burden of proof. The growing versality of robots justifies shifting the burden of proof from the plaintiff, assuming that the robot was obligated to act, to the person responsible for the robot, assuming that the robot was not obliged to act. In this case, however, other problems arise. Is the robot obliged to respond to the risk of harming humanity in general, or to the risk of harming any specific person present in a particular position or relationship with the robot? Or should the catalogue of protected persons be limited by specific criteria, such as spatial proximity, legal relationships or age? It would be unreasonable to expect a robot to react to any possible danger to humanity, or an individual, and hence, to create a robot and maintain it in such a state that it has such a capability. Such a legal obligation should therefore be excluded by the principle of *impossibilum nulla est obligatio*. If this is the case, the problem of setting the criteria becomes inevitable.

However, putting aside these doubts for the moment and assuming that this responsibility is to be a legal one, the question of its nature arises. Should it be based on criminal responsibility, civil responsibility, contractual liability or tort liability, or would some new form of responsibility enacted in the future be required?

Furthermore, it should be remembered that the responsibility placed on human beings will be, to some degree, dependent on the future legal status of robots; this will be awarded by new legislation regarding the so-called "robot law" currently being drafted. The results of this legislation will depend on whether the robots remain as *things*, or if all or some of them are recognised as being endowed with legal personality, at least within some scope of affairs. In the latter, it is possible to imagine that at least solidary or subsidiary liability will be awarded.

Should the development of AI continue, it will eventually give rise to a strong AI, or maybe some intermediate form. The natural consequence of such evolution would be, contrary to various emotional protests,[20] the endowment of robots with even residual legal personality, as demanded by Resolution 2017. The possibility that autonomous decisions may be made by AI inevitably raises the question of whether these decisions and any consequent actions might be inconsistent with the terms of reference assumed by its creators. This problem will be most evident in such cases when AI makes decisions and valuations which are not *a priori* feasible or describable. In such cases, it will be reasonable to make the robot itself the addressee—and strictly speaking, the second addressee—of Asimov's Laws, according to the

[20] See Open Letter to the European Commission. Artificial Intelligence and Robotics, 2019. http://www.robotics-openletter.eu, last access on the 4th of August 2022.

primary concept. Although the creator should design the AI according to certain ethical and legal assumptions, and within certain limits, it is the robot who will make decisions based on them, and hence the assumptions should be consistent. Therefore, breaching the terms of reference must imply some consequences in the form of some kind of responsibility or liability. This may be realised in the form of civil liability, or as a specific personal liability consisting of eliminating this type of AI from the market.

If the developed strong AI is endowed with some form of legal personality, the robot it operates should consequently be required to act at least in such a way that could be legally demanded from a human being, within the scope of the Zeroth and the First Law. If, for example, the legal rules require a human being to actively help a person whose life is in danger, it is all the more justified that the robot is burdened with the same obligation. However, in legal systems where no such general obligation is imposed on people, placing it on robots is a rather idealistic requirement. For example, if the robot were able to help a person on another continent, should it do so, according to the literal interpretation of Asimov's Law? Or instead, should a superintelligent robot (AI system) act for humanity's benefit, understood according to its own criteria? Indeed, it is a common knowledge that for humanity, the greatest hazard is man himself.

At first sight it is obvious that each of Asimov's Laws protects different set of goods and subjects, or does so to varying degrees:

1. The function of the Zeroth Law is to protect humanity. It relates to the situations and relationships between the person responsible for a robot and all human beings, or groups of human beings, with humanity itself as the largest group.
2. The First Law serves to protect human life and human health in its broadly-understood sense, and to ensure human well-being. However, some people may invoke, on account of legal policy, the broad concept of harm and assume that the First Law protects also the property rights of human beings. In our opinion, this would be an overinterpretation of Asimov's Laws, whose original function is to ensure physical safety for human beings and nothing more. It should be remembered that in Asimov's Laws, the words "harm" or "injury" are used in the commonly-understood sense rather than the legal one, and certainly not in the meaning specific for a concrete legal system (e.g. German, French or Italian). Moreover, the problem would then arise of why only the property of human beings should be protected, and not that belonging to juridical persons, even though it is obvious that Asimov's Laws do not relate to juridical persons at all. Asimov's First Law relates to the situations and relationships between a person responsible for a robot and each potentially harmed human being.
3. The Second Law serves to protect the expectations addressed to a robot by a human being. It relates to the function which the robot is to perform. However, it cannot be probably assumed that all such expectations of all people should be protected, not least because such expectations could be conflicting with others. It seems that this Law should protect only the interest of the human beings who

have, because of this or another title, the right to command a robot or to issue instructions.
4. The Third Law serves, at least today, to protect property which belongs to some legal subject, be it a human being, physical person, juridical person and so on; part of this property is the robot itself. If robots become independent to such an extent that they could be recognised within a particular scope as non-things, they would still be protected by this law; for example, if they were endowed with a status similar to that of animals, which in some legal systems are not recognised as either legal subjects or as things, or if they were recognised as legal subjects. The Third Law would also protect the existence and well-being of robots, even if it were endangered by the influence of some unauthorised human beings: for example, a person who had hacked it and commanded it to self-destruct.

As such differences exist between these Asimov's Laws, it is necessary to clearly distinguish different types of the legal responsibility for their observance, even if it is ascribed to human beings:

1. Obviously, for the Zeroth Law, the methods of regulation characteristic for private law are not suitable. Civil law does not protect "humanity" as a whole and "humanity" cannot claim redress for damage by means of civil law instruments. As the construction of civil law makes sanctioning the robot's action towards humanity rather doubtful, the Zeroth Law is simply not practical in this context. On the other hand, because of the considerable risk associated with breaching this Law, the methods associated with criminal law and administrative law may be more adequate. Such responsibility should be unlimited in time, though it can be transferred in sequence to the subjects controlling a robot or may be shared by them on the basis of various rules. The producer or designer may be held initially responsible, followed by the software or hardware service agent, data manager, operator or owner. In such cases, the robot registration system and accompanying compulsory insurance may be of some importance. With the growing risks associated with the use of autonomous robots, i.e. ownerless robots operating without being connected to anybody, there will be a growing need to enact special guarantee funds intended to pay for any damages they may incur.[21]
2. For the First Law, the two main legal regimens, criminal and civil, would be eminently suitable. Both regard violation of goods or values as the reason for imposing a sanction; and the justification, in common language, may be called *a harm* or *an injury*. So, if it is assumed that the basic function of civil liability is to redress damage, harm or injury, mentioned in the Zeroth and First Laws, can only mean "inflicting damage" in the context of civil law. Hence, this kind of harm, be it toward an individual human being or toward humanity as a whole, and irrelevant in regular civil law, has to remain irrelevant in any new civil law affecting robots. It is difficult to imagine the operation of a normative system in which actions that are treated as illegal and "harming" when performed by a

[21] This way of thinking presents Proposal 2021.

machine would be permissible for human beings, assuming that they do not cause harm within a traditional conceptualisation. Any attempt to make such a differentiation between humans and machines will be unsuccessful *a priori*, as damage, be it property or personal, must be related to the person upon which a damage was inflicted and not to the perpetrator or her behaviour. If no damage was inflicted on a person, in the sense of the valid principle of civil law, it is not possible to insist that he was harmed simply because the action was performed by a robot.

From this perspective, it appears obvious that the principle "do not harm", if interpreted on the basis of civil law, is strictly connected with the general rules of tort liability, though its true content can be perceived only in relation to the concrete legal order. Adducing general rules would necessarily result in the scope of protection becoming differentiated in response to the internal or national legal order; this would undoubtedly entail different interpretations of Asimov's Laws at every turn. If, for example, the given legal order does not accept compensation for so-called moral damages or indirect damages, or imposes limits on the amount of compensation, the normative content of Asimov's Laws will differ to those enforced in places where such compensation is acceptable or no limits are present. Certainly, it is possible, and within the European Union achievable and even desirable, to construct rules which independently determine both the content of the concept of damage and the conditions for its compensation, as well as the standard of compensation when the perpetrator is a robot. However, if such a regulation were to be introduced, it would not have been a consequence of Asimov's Laws in any way.

Of course, *a harm* or *damage*, on the basis of a concrete legal system, may theoretically be understood as any worsening of the property or well-being of an individual (Asimov's First Law) or even of humanity (the Zeroth Law). But let's take a simple case which happens very often nowadays. It is not uncommon for stock market transactions to be made by an AI, in non-physical form, or similarly, games such as poker to be played for money. In both cases, if the AI wins, it lowers the status of the property of its competitors or stock market contractors, and can potentially negatively impact their well-being. Alternatively, a robot, i.e. an AI in physical form, that cleans a depot may work so well that its predecessor, a human employee, is dismissed and his financial situation worsens. In both cases, a *harm* is permissible, because it is qualified on the basis of an appraisal of behaviour according to the rules of the given legal order, i.e. from the perspective of its legality or illegality. In this case, the simple fact that an action causes, in the broad sense of the word, the worsening of someone's condition (*a harm*) is not enough to recognise it as illicit. It is necessary to qualify this harm through the lens of a given legal order, which in many borderline cases, typically relies on fuzzy valuations, extra-legal axiological systems, the rules of ethics and suchlike. Accordingly, the question can be posed whether the Zeroth and the First Laws have any other content, and whether they should be interpreted differently than through the moral imperative of not harming.

If, in spite of the above caveats, Asimov's First Law must be forced into civil law, then a new liability is required that is analogous to existing forms: general

liability for property and personal damage, liability for damage inflicted by unsafe product, liability for damage inflicted by animals, liability for damage caused by mechanical means of transportation powered by the forces of nature. This liability should be unlimited in time, though it can be transferred in sequence onto the subjects controlling a robot, or shared by them on the basis of various rules, such as the solidary liability of producer and operator, the guarantee liability of insurance company, or the subsidiary liability of a guarantee fund. The producer or designer could initially be held liable, followed by the software and hardware service man, operator and owner.

In the light of the above analysis, it should be assumed that the First Law must ultimately be translated into the rules elaborated within the civil law doctrine. The prohibition of harm towards humans is nothing more than the old principle of *neminem laedere* and the principle of liability for inflicting damage. Such damage may be inflicted on the personal goods of an individual or his property, including other robots. The conditions of this liability presumably should disregard the subjective elements, such as fault, as given in the case of liability for damage inflicted by an unsafe product, or liability for damage caused by a mechanical means of transportation powered by the forces of nature; however, this is not entirely evident in the case of an autonomous robot, especially one controlled by a strong or nearly strong AI. There are no reasons to assume *a priori* that any institution of liability for damages inflicted by a robot should be absolute and independent of such circumstances as fault or the injured person's fault (attributability). In particular, as is the case with the concept of the cause-effect relationship, the concept of fault will require re-evaluation within the civil law framework as robots take a growing and increasingly active role in both social life and legal practice.

The First Law cannot serve as an instruction for civil law that can be used to determine the subjects against which a robot cannot act. It is quite obvious that redress for damage can also be claimed by various subjects of civil law other than human beings, such as a juridical person. When a robot is endowed with some kind of personality, it could also be regarded as a victim of some form of damage. It is entirely possible that litigation may take place between robots at some time in the future.

3. For the implementation of the Second Law, the most adequate form would be civil liability for improper performance of obligation or liability for warranty or guarantee. It is difficult to justify any other forms of liability when it is only the robot's behaviour, consisting of disobedience to a human command, that is in question, and no damage is caused. However, there is room for discussion regarding whether such a liability should be limited in time and if so, its duration. Certainly, technically it would be very difficult to realise the demand that the Second Law must not collide with the First Law; in addition, a collision between the obligations to obey orders issued by humans and the obligation to avoid harming them may cause even problems from the perspective of civil liability. Avoiding this situation requires the robot to estimate the degree of harm, or of well-being, arising from its actions, or to evaluate which commands should be

obeyed and which should not; on a more complex level, the robot could be required to compare the potential damage which would be caused to his disposer by not obeying a command with the potential general harm which would be caused if the command were obeyed.

Such problems appear simply insoluble. If it is recognised that the enacted commands realise some interest by the authorised disposer, it should also be acknowledged that disobeying the command may, or even certainly, result in interference with this interest and some form of damage to the disposer. It is certainly possible to create some systems that can prevent certain or significant external damages, such as killing or injuring a human being; however, when the damage is less obvious for an external observer, such as transgressing the limits of the freedom of speech or damage in the form of *lucrum cessans*, it seems impossible to make a system which could evaluate whether commands should be obeyed and to determine whether damage resulting from a course of action may be acceptable. As an infinite number of consequences is possible for each action, especially when considering indirect ones, the analysis of potential damages inflicted by a robot while obeying a command should be limited in some way.

4. For the Third Law, only analogues to civil liability for warranty or guarantee would be adequate; however, it is debatable whether such a liability should be unlimited in time, or how long it should last. The problems described while relating to the Second Law are even more visible here. A robot would have to make a multidirectional evaluation on the potential damages which could be caused by its action or non-action. The application of the Third Law would force the conclusion that a robot could not undertake any efficient defensive actions against people, as most of these actions would surely result in harm to a human being.

In practice, it is possible to approach the Third Law in two ways. The first would be to assume that this law is not applicable at all to relations between a robot and a human being, i.e. that a robot could not defend itself against dangers originating from human beings. Such a conception appears to be an effective way to assuage the fears regarding the rise of the machines. The second would be to allow some kind of self-defence in this relationship; however, this apparently rational approach imperils human dignity, because a human being could be put in a hierarchically worse position than a machine. From a more long-term perspective, it is possible that the physical and intellectual strength of robots would surpass that of humans, resulting in humans eventually having to submit their will to the AI. It would therefore be more prudent to consistently treat a robot as a thing unable to enact its will, and forbid it from defending itself against a human being. However, if some form of personification of AI were to be accepted, and this state of affairs looks inevitable as we approach strong or even self-conscious AI, it also appears inevitable that it will have the right to defend itself. In this case, it would be difficult indeed to invent a reason why a robot should be limited in exercising the Third Law, i.e. why it should sacrifice its personality and interests for those of a human being.

It should be remembered that in the case of the Laws being breached by a robot under the influence of an unauthorised party, i.e. someone who "hacked" the robot, thus making him the direct perpetrator of the breach, it is the hacker who would be liable, civilly or criminally, perhaps together with the person authorised to control the robot for negligence in supervision. However, if robots were endowed with some scope of legal personality, they may bear autonomous, solidary or subsidiary liability, or maybe even criminal liability: this would be analogous to the criminal liability borne by juridical persons (collective entities) in certain legal systems. Of course, such endowment with legal personality, as explained above, would cause conflict with the Second Law, which demands obedience toward human beings. The presence of strong autonomy, resulting from a high level of multidimensional intelligence, combined with a strong ability to influence physical reality, would make it necessary to subjectivise their liability: it would force the robot's behaviour to be evaluated not only through the lens of compliance with the legal rules, but also, as in the case of human beings, on the basis of some subjective premises or concepts, such as fault.[22]

After Resolution 2017 was issued, a wide public discussion took place. As a result of various reservations in Resolution 2020, the European Parliament gave up the idea of awarding legal status or direct responsibility to robots. Nevertheless, in the Annex, the European Parliament included principles and aims intended to guide regulation regarding the liability for operation of Artificial Intelligence systems (A.), and provided a proposal for a European Parliament and Council regulation regarding liability for the operation of Artificial Intelligence systems (B.). Both parts of the Annex are based on the clear assumption that the liable entity cannot be an AI system itself, but natural or legal persons; these can either be the producer, when liability is based on PLD, or the operator, understood as the frontend or backend operator (Annex, B. Article 3 d–f). In the latter case, the operator is the person who is controlling the risk, and is comparable to an owner of a car (Introduction 10). The European Parliament also underlines that

> [a]ny liability framework should strive to instil confidence in the safety, reliability and consistency of products and services, including emerging digital technologies [. . .] in order to strike a balance between efficiently protecting potential victims of harm or damage and at the same time providing enough leeway to make the development of new technologies, products or services possible. Resolution 2020 (Annex, B. (1))

The European Parliament goes on to explain that it:

> Believes that there is no need for a complete revision of the well-functioning liability regimes [. . .]
>
> The rise of AI, however, presents a significant challenge for the existing liability frameworks. Using AI-systems in our daily life will lead to situations in which their opacity ("black box" element) and the multitude of actors who intervene in their life- cycle make it extremely expensive or even impossible to identify who was in control of the risk of using the AI-system in question or which code or input caused the harmful operation. That

[22] The above ideas about the Asimov's Law were presented in Księżak and Wojtczak (2020).

> difficulty is compounded by the connectivity between an AI-system and other AI-systems and non-AI-systems, by its dependency on external data, by its vulnerability to cybersecurity breaches, as well as by the increasing autonomy of AI- systems triggered by machine-learning and deep-learning capabilities. In addition to these complex features and potential vulnerabilities, AI-systems could also be used to cause severe harm, such as compromising human dignity and European values and freedoms, by tracking individuals against their will, by introducing social credit systems, by taking biased decisions in matters of health insurance, credit provision, court orders, recruitment or employment or by constructing lethal autonomous weapon systems. Resolution 2020 point 5 and Annex. B. (3))

According to the Resolution 2020 Annex. B. Article 3, the operator of a high-risk AI system should be strictly liable for any harm and damage that was caused by any physical or virtual activity, device or process driven by that AI-system (section 1) and the operator could not be exonerated from liability by arguing that he or she acted with due diligence, or that the harm or damage was caused by an autonomous activity, device or process driven by AI-system; however, the operator should not be held liable if the harm or damage had been caused by force majeure (section 3). The compensation should be limited as to its amount and extent (Articles 5 & 6).

According to the same document (Annex. B. Article 8) the operator of an AI system, although not a high-risk one, should be subject to fault-based liability for any harm or damage that was caused by a physical or virtual activity, device or process driven by the AI system;[23] however, in such cases, the operator should be presumed to be at fault (point [17] sentence 3). The operator, however, can be exonerated on the basis of one of two grounds: (a) the AI system was activated without his or her knowledge and that all reasonable and necessary measures to avoid such activation outside of the operator's control were taken, or (b) due diligence was observed by performing all the following actions: selecting a suitable AI-system for the right task and skills, putting the AI-system duly into operation, monitoring the activities and maintaining its operational reliability by regularly installing all available updates. Similarly to the case of a high-risk AI system, the operator could not be exonerated from liability by arguing that the harm or damage was caused by an autonomous activity, device or process driven by the AI system, although the operator should not be held liable if the harm or damage had been caused by force majeure. The proposal includes also the institutions of contributory negligence of an affected person or of any person for whom the affected person is responsible (Article 10), joint and several liability when more than one operator was using the AI system (Article 11) and recourse for compensation (Article 12).

This thread seen in the draft AI legislation also seems to run through later proposals by EU bodies. It is particularly clearly visible in the Proposal 2021, where the division between high-risk and non-high-risk AI-systems is the starting point; the document proposes criteria for classifying AI systems as high risk and

[23] However, point (17) of Annex B includes an additional condition: "unless stricter national laws and consumer protection legislation is in force. The national laws of the Member States, including any relevant jurisprudence, with regard to the amount and extent of the compensation, as well as the limitation period, should continue to apply."

strongly promotes the creation of a register for high-risk AI systems. The demands, including formal ones, towards high-risk AI-systems are strong, while the ones towards non-high risk are quite weak. Such a situation should be criticized in the context of the regime of civil liability, which according to European Parliament is to be different for these two kinds of AI.

First, the division is entirely conventional and to certain degree, arbitrary. Resolution 2020 gives the following general definition of *high-risk*:

> 'high risk' means a significant potential in an autonomously operating AI-system to cause harm or damage to one or more persons in a manner that is random and goes beyond what can reasonably be expected; the significance of the potential depends on the interplay between the severity of possible harm or damage, the degree of autonomy of decision-making, the likelihood that the risk materializes and the manner and the context in which the AI-system is being used [. . .]

In contrast, the later, and more influential, Proposal 2021 indicates:

> **Article 6**
> *Classification rules for high-risk AI systems*
>
> 1. Irrespective of whether an AI system is placed on the market or put into service independently from the products referred to in points (a) and (b), that AI system shall be considered high-risk where both of the following conditions are fulfilled:
>   (a) the AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonisation legislation listed in Annex II;
>   (b) the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonisation legislation listed in Annex II.
> 2. In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall also be considered high-risk.

Annex II includes a List of Union harmonisation legislation based on the New Legislative Framework (section A) and a List of other Union harmonisation legislation (section B.), while Annex III includes a list of particular areas of activity and the types of AI systems acting within them; these may be updated by the Commission in accordance with the determined criteria given in Article 7. It can be seen that the classification is a rigid one which fails to consider many of the intermediate categories that exist in practice. As such, the consequences of attributing a concrete AI system to one of the two classes would be very severe and may cause unjust results.

Secondly, the rules propose that high-risk AI systems should be more strictly controlled than the lower-risk forms. They will be tested, registered, monitored and documented in a special way and by different legally-determined bodies. Furthermore, the regime of liability would prevent the operator being exonerated on the

basis of various due-diligence arguments, such as selecting a suitable AI-system for the right task and skills, putting the AI-system duly into operation, monitoring its activities and maintaining operational reliability by regularly installing all available updates. In this case, while the state and the law would both control the AI system, the risk associated with its use would be borne by the operator, which appears unjust.

Moreover, developing such a system of liability which explicitly ignores the fact that some harms or damages may be caused by autonomous activities, devices or processes driven by the AI system, may only represent a short-term strategy. Firstly, such situations, where the harm or damage is caused by autonomous activity, would become increasingly frequent, often in a short period of time. Secondly, it would increase the cost and complexity of identifying all liable entities and settling the so-called liability chain would become more and more complicated and expensive. Thirdly, excluding the possibility to escape liability by the operator when his role in the process of invoking the harm or damage is negligible, would probably seem unjust for the general public. Hence, at least in certain cases, the liability for harm or damage should be attributed directly to the AI system, which in these circumstances, should be treated as the subject of the law. Of course, in most of the cases such responsibility should be based on strict liability.

Such a solution appears a rational one, especially considering the proposal included in the Resolution 2020, which postulates the implementation of mandatory insurance against civil liability for damages caused by high-risk AI systems. It also proposes the creation of a special compensation fund to cover exceptional cases, such as when harm or damage is caused by an AI system which is not yet classified as high risk and thus, is not yet insured, (Annex. B. [21–22]). Such measures may allay the fears of those opposing legal subjectivity for AIs, i.e. that making AI liable may deprive the victim of the possibility of obtaining redress. However, it should be taken into account that further European acts, i.e. especially Proposal ALD 2022 (p. 6), assume introducing such measures later, only at the second stage of the process of adapting non-contractual civil liability rules to AI.

11.2 Basic Concepts

To reflect on these issues, we should return to certain basic concepts. However, the aim of this section is only to provide a general, rough outline sufficient for the topic. It is not the intention to perform a detailed analysis of all the rich advantages associated with the civil law doctrine in different countries. It will focus on the fact that some common standards or constructions exist in Western legal culture, and that these are described in various documents such as DCFR or PETL which attempt to collect so-called model rules.

First, it is necessary to bear in mind that in the domain of civil law, obligations may be contractual or non contractual (extracontractual). While contractual liability derives from the terms of contracts and applicable jurisdictional clauses, non-contractual liability arises from the rules imposed by the law for the protection of certain legally-valuable rights.

A fundamental concept within the field of liability is damage. It is important to note that not every form of harm, loss or damage entails liability, but only legally-relevant damage. For instance, in the context of contractual liability, DCFR defines such a loss in the following way:

III. – 3:701: Right to damages

(1) The creditor is entitled to damages for loss caused by the debtor's nonperformance of an obligation, unless the non-performance is excused.
(2) The loss for which damages are recoverable includes future loss which is reasonably likely to occur.
(3) "Loss" includes economic and non-economic loss. "Economic loss" includes loss of income or profit, burdens incurred and a reduction in the value of property. "Non-economic loss" includes pain and suffering and impairment of the quality of life.

III. – 3:702: General measure of damages

The general measure of damages for loss caused by non-performance of an obligation is such sum as will put the creditor as nearly as possible into the position in which the creditor would have been if the obligation had been duly performed. Such damages cover loss which the creditor has suffered and gain of which the creditor has been deprived.

III. – 3:703: Foreseeability

The debtor in an obligation which arises from a contract or other juridical act is liable only for loss which the debtor foresaw or could reasonably be expected to have foreseen at the time when the obligation was incurred as a likely result of the non-performance, unless the non-performance was intentional, reckless or grossly negligent.

While in the case of non-contractual liability, DCFR defines legally-relevant damage as:

VI. – 2:101: Meaning of legally relevant damage

(1) Loss, whether economic or non-economic, or injury is legally relevant damage if:
  (a) one of the following rules of this Chapter so provides;
  (b) the loss or injury results from a violation of a right otherwise conferred by the law; or
  (c) the loss or injury results from a violation of an interest worthy of legal protection.
(2) In any case covered only by sub-paragraphs (b) or (c) of paragraph (1) loss or injury constitutes legally relevant damage only if it would be fair and reasonable for there to be a right to reparation or prevention, as the case may be, under VI. – 1:101 (Basic rule) or VI. – 1:102 (Prevention).

(3) In considering whether it would be fair and reasonable for there to be a right to reparation or prevention regard is to be had to the ground of accountability, to the nature and proximity of the damage or impending damage, to the reasonable expectations of the person who suffers or would suffer the damage, and to considerations of public policy.
(4) In this Book:
    (a) economic loss includes loss of income or profit, burdens incurred and a reduction in the value of property;
    (b) non-economic loss includes pain and suffering and impairment of the quality of life.

The document gives various particular instances of legally-relevant damage, including the following: personal injury to a natural person's body or health and consequential loss (VI. – 2:201), loss suffered by third persons as a result of another's personal injury or death (VI. – 2:202), infringement of dignity, liberty and privacy (VI. – 2:203), loss upon communication of incorrect information about another (VI. – 2:204), loss upon breach of confidence (VI. – 2:205), loss upon infringement of property or lawful possession (VI. – 2:206), loss upon reliance on incorrect advice or information (VI. – 2:207), loss upon unlawful impairment of business (VI. – 2:208), burdens incurred by the state upon environmental impairment (VI. – 2:209), loss upon fraudulent misrepresentation (VI. – 2:210), loss upon inducement of non-performance of obligation (VI. – 2:211).

In turn, PETL defines damage in the following way, within tort liability:

Art. 2:101. Recoverable damage
Damage requires material or immaterial harm to a legally protected interest.

Art. 2:102. Protected interests
(1) The scope of protection of an interest depends on its nature; the higher its value, the precision of its definition and its obviousness, the more extensive is its protection.
(2) Life, bodily or mental integrity, human dignity and liberty enjoy the most extensive protection.
(3) Extensive protection is granted to property rights, including those in intangible property.
(4) Protection of pure economic interests or contractual relationships may be more limited in scope. In such cases, due regard must be had especially to the proximity between the actor and the endangered person, or to the fact that the actor is aware of the fact that he will cause damage even though his interests are necessarily valued lower than those of the victim.
(5) The scope of protection may also be affected by the nature of liability, so that an interest may receive more extensive protection against intentional harm than in other cases.

(6) In determining the scope of protection, the interests of the actor, especially in liberty of action and in exercising his rights, as well as public interests also have to be taken into consideration.

Another indispensable element of civil liability is the causation link between the action of the liable entity and the damage, or other way of accountability (e.g. probability or correlation). Although these issues may be based on different principles, they share some basic ones and some others which regulate specific cases.

A general rule given by DCFR is:

VI. – 4:101: General rule

(1) A person causes legally relevant damage to another if the damage is to be regarded as a consequence of that person's conduct or the source of danger for which that person is responsible.
(2) In cases of personal injury or death the injured person's predisposition with respect to the type or extent of the injury sustained is to be disregarded.

VI. – 4:102: Collaboration

A person who participates with, instigates or materially assists another in causing legally relevant damage is to be regarded as causing that damage.

While a basic rule given by PETL is:

Art. 3:101. Conditio sine qua non

An activity or conduct (hereafter: activity) is a cause of the victim's damage if, in the absence of the activity, the damage would not have occurred.

The specific rules are those regarding alternative causes as noted in DCFR:

VI. – 4:103: Alternative causes

Where legally relevant damage may have been caused by any one or more of a number of occurrences for which different persons are accountable and it is established that the damage was caused by one of these occurrences but not which one, each person who is accountable for any of the occurrences is rebuttably presumed to have caused that damage.

And in PETL:

Art. 3:103. Alternative causes

(1) In case of multiple activities, where each of them alone would have been sufficient to cause the damage, but it remains uncertain which one in fact caused it, each activity is regarded as a cause to the extent corresponding to the likelihood that it may have caused the victim's damage.
(2) If, in case of multiple victims, it remains uncertain whether a particular victim's damage has been caused by an activity, while it is likely that it did

not cause the damage of all victims, the activity is regarded as a cause of the damage suffered by all victims in proportion to the likelihood that it may have caused the damage of a particular victim.

Other specific rules are listed in PETL:

Art. 3:104. Potential causes
(1) If an activity has definitely and irreversibly led the victim to suffer damage, a subsequent activity which alone would have caused the same damage is to be disregarded.
(2) A subsequent activity is nevertheless taken into consideration if it has led to additional or aggravated damage.
(3) If the first activity has caused continuing damage and the subsequent activity later on also would have caused it, both activities are regarded as a cause of that continuing damage from that time on.

Art. 3:105. Uncertain partial causation
In the case of multiple activities, when it is certain that none of them has caused the entire damage or any determinable part thereof, those that are likely to have [minimally] contributed to the damage are presumed to have caused equal shares thereof.

Art. 3:106. Uncertain causes within the victim's sphere
The victim has to bear his loss to the extent corresponding to the likelihood that it may have been caused by an activity, occurrence or other circumstance within his own sphere.

The problem of damage and causation is strictly connected to the mechanisms of burden of proof. Although generally the burden of proof is attributed to the entity that asserts the legal consequences arising from the fact, some exceptions to this principle exist. These exceptions typically arise as a result of some inequality between the parties to the relationship, or the willingness of the law maker to protect some legally-relevant values. One example can be seen in PETL:

Art. 4:201. Reversal of the burden of proving fault in general
(1) The burden of proving fault may be reversed in light of the gravity of the danger presented by the activity.
(2) The gravity of the danger is determined according to the seriousness of possible damage in such cases as well as the likelihood that such damage might actually occur.

There are at least three regimes of liability in civil law: one based on fault, another based on strict liability (for things, for dangerous activities and for animals) and another on liability for others (vicarious liability). While the most widely-applicable

form of liability in Western countries is that based on fault, strict liability is gaining precedence. Although in most legal systems still the fault liability is a principle and the strict liability is an exception, more and more types of cases are being submitted to strict liability (by legislation).[24] Also the majority of proposals on legal regulation of emergent technologies acknowledge strict lability as the most applicable to harms caused by AI. However, the fault regime remains a key influence in some specific domains and situations, as discussed below, and cannot be ignored when examining the forthcoming changes in civil law demanded by the advance of emerging technologies and AI. This necessity of starting from a minimally invasive approach was noticed by the European Commission who, preparing the rules for the liability for AI, proposed two separate acts: Proposal DLDP 2022, embracing the strict liability regime, and Proposal ALD 2022, embracing the fault liability regime. Furthermore, the Commission declared in Proposal ALD (p. 6) that it will take into account further jurisprudential developments including for situations where strict liability would be more appropriate, later, at the second stage of the legislative process, after re-assesing the need for more stringent or extensive measures than introduced in this proposal.

As noted in DCRF:

VI. – 1:101: Basic rule

(1) A person who suffers legally relevant damage has a right to reparation from a person who caused the damage intentionally or negligently or is otherwise accountable for the causation of the damage.

(2) Where a person has not caused legally relevant damage intentionally or negligently that person is accountable for the causation of legally relevant damage only if Chapter 3 so provides.

and in PETL:

TITLE III. Bases of Liability

Chapter 4. Liability based on fault

Section 1. Conditions of liability based on fault

Art. 4:101. Fault

A person is liable on the basis of fault for intentional or negligent violation of the required standard of conduct.

Art. 4:102. Required standard of conduct

(1) The required standard of conduct is that of the reasonable person in the circumstances, and depends, in particular, on the nature and value of the protected interest involved, the dangerousness of the activity, the expertise to

[24] A very clear juxtaposition of current national liability frameworks in EU is included in Evas (2020).

be expected of a person carrying it on, the foreseeability of the damage, the relationship of proximity or special reliance between those involved, as well as the availability and the costs of precautionary or alternative methods.

(2) The above standard may be adjusted when due to age, mental or physical disability or due to extraordinary circumstances the person cannot be expected to conform to it.

(3) Rules which prescribe or forbid certain conduct have to be considered when establishing the required standard of conduct.

Hence, for fault-based liability, in addition to the obvious conditions of legally-relevant damage and the causation of damage,[25] it is also necessary to include the violation of the required standard of conduct (or standard of care), or the attitude of the perpetrator of the damage towards the deed.

Standard of conduct (or standard of care, or due diligence) is an objective element of the concept of fault accepted in civil law. Instead of being modelled on the behaviour of any concrete person, its demands are based on the actions of a hypothetical reasonable person acting in a certain context. The elements of standard of conduct are also significant, being the means for realisation of values protected by law. Hence, standard of conduct may depend on various factors, such as the nature of the activity of a person obliged by this standard (e.g. commercial vs non-commercial), the type of activity (e.g. dangerous vs non-dangerous), the relationship linking the parties (e.g. one based on equality vs one based on subordination or care) and the capabilities and qualifications of the person obliged to keep the standard (over certain age vs under certain age; professional vs non-professional), among others.

In contrast, the attitude of the perpetrator towards the damaging action is a less objective element of the concept of fault, since to determine this, it is necessary to examine whether the deed arose intentionally or negligently.

It is difficult to apply the existing concept of fault to entities other than human beings or collective entities. Particularly, when "translating" this concept to accommodate liability by AI, two main problems arise: firstly, the problem of what standard of conduct (standard of care) should be required from AI and how it can be reconciled with the standard imposed on human beings, and secondly, how to determine the attitude of the AI towards its actions when it lacks consciousness and has no psychological or mental experiences.

[25] The problems of causation appearing when albeit perpetrator of harm is AI and some proposals of solving these problems are presented in Wojtczak and Księżak (2021).

11.3 Legally-Relevant Damage Caused by AI

Resolution 2020 point 19 indicates the values whose infringement should be protected by the civil liability of AI:

> [. . .] the proposed Regulation should cover violations of the important legally protected rights to life, health, physical integrity and property [. . .]

and gives a recommendation as to the extent of legally-relevant damage potentially caused by AI, especially whether immaterial harm should be included in this category:

> is of the opinion that the proposed Regulation should also incorporate significant immaterial harm that results in a verifiable economic loss above a threshold harmonised in Union liability law, that balances the access to justice of affected persons and the interests of other involved persons; urges the Commission to re-evaluate and to align the thresholds for damages in Union law; is of the opinion that the Commission should analyse in depth the legal traditions in all Member States and their existing national laws that grant compensation for immaterial harm, in order to evaluate if the inclusion of immaterial harm in AI-specific legislative acts is necessary and if it contradicts the existing Union legal framework or undermines the national law of the Member States [. . .]

As a result, the European Parliament's draft of the regulation included in Annex B. Article 2.1 indicates that this liability should concern harm or damage to the life, health, physical integrity of a natural person, to the property of a natural or legal person, or significant immaterial harm resulting in a verifiable economic loss. The compensation should be limited by amount (Annex B. Article 6).

It is important to note the emphasis placed by the proposal on immaterial loss. It has been included despite the knowledge shared in European Parliament that some EU member legal systems are reluctant to endorse monetary damages for immaterial losses.

In turn, Annex II and Annex III of the Proposal 2021, concerning high-risk AI systems, include a list of the products and areas which, according to the Commission, appear most vulnerable to damages: machinery, toys, recreational craft and personal watercraft, lifts, equipment and protective systems intended for use in potentially explosive atmospheres, radio equipment, pressure equipment, cable installation, personal protective equipment, appliances burning gaseous fuels, medical devices, in vitro diagnostic medical devices, civil aviation security, two- or three-wheel vehicles and quadricycles, agriculture and forestry vehicles, marine equipment, rail system, motor vehicles and their trailers, unmanned aircraft, biometric identification and categorisation of natural persons; management and operation of critical infrastructure; education and vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private services and public services and benefits; law enforcement; migration, asylum and border control management; administration of justice and democratic processes.

This list seems strange and rather chaotic, although it may provide an insight into the preferences of the Commission. In particular, confirming the conclusions given

by EP PRCLR 2020, it seems that they are most preoccupied not by economic losses, but by infringements of fundamental rights and the potential overuse of advantages provided by AI. This is, to some extent, justified by the current and commonly paradigmatic (prototypic) and *prima facie* character of these domains as potential "places" where AI may be dangerous. However, it may be suspected that this list will be changed in the future.

It is also important that Proposal DPDL 2022 (Article 4 point 6c) broaded the definition of damage to encompass the material losses resulting from loss or coruption of data, when they not used for professional purposes, however in Proposal ALD (p. 11) it is declared that "the measures provided in this Directive [. . .] reflect an approach that does not touch on the definition of fundamental concepts like 'fault' or damage', given that the meaning of those concepts varies considerably accross the Member States".

11.4 Causation

The problem of establishing the causative link between action and damage is an especially important one, as causation is a condition of liability in both strict-liability and fault-based liability. This issue was raised quite early in the debate on AI law as a consequence of the wide and long-time dissemination of knowledge about the opacity and unexplainability of AI systems. Because of the commonly-made analogy between the computer and the human mind, any difficulties encountered within this area are often compared with problems specific to medical science, particularly in neuroscience. As a result, the term "black box" has been borrowed from behavioural psychology, where it is used to describe a human mind:[26] AI systems are commonly seen as "black boxes", although such a position is questionable.

Resolution 2020 motive H explains:

> whereas certain AI-systems present significant legal challenges for the existing liability framework and could lead to situations in which their opacity could make it extremely expensive or even impossible to identify who was in control of the risk associated with the AI-system, or which code, input or data have ultimately caused the harmful operation; whereas this factor could make it harder to identify the link between harm or damage and the behaviour causing it, with the result that victims might not receive adequate compensation [. . .]

and then in point 7:

> is of the opinion that the opacity, connectivity and autonomy of AI-systems could make it in practice very difficult or even impossible to trace back specific harmful actions of AI-systems to specific human input or to decisions in the design [. . .]

Of course, many European reports and documents demand to promote and realize AI transparency; for example, Proposal 2021 (point 47) notes:

[26] Skinner (1969), p. 282; Nathan (2021), pp. 61–67.

> To address the opacity that may make certain AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for high-risk AI systems. Users should be able to interpret the system output and use it appropriately. [. . .]

However, although achieving transparency or explainability is theoretically possible[27] it may be difficult in practice[28] for various technical (the complexity of the systems), legal (the confidential information involved in the action of AI), economic (the expense of ensuring explainability) and cognitive reasons (the need for high-performance AI specialists to perceive and understand the action of AI), among others. As such, some special remedies are needed to make the causation concept feasible.

For instance, the report from the Expert Group on Liability and New Technologies—New Technologies Formation, Liability for Artificial Intelligence and Other Emerging Digital Technologies (2019) proposes:

> [20] There should be a duty on producers to equip technology with means of recording information about the operation of the technology (logging by design), if such information is typically essential for establishing whether a risk of the technology materialized, and if logging is appropriate and proportionate, taking into account, in particular, the technical feasibility and the costs of logging, the availability of alternative means of gathering such information, the type and magnitude of the risks posed by the technology, and any adverse implications logging may have on the rights of others.
>
> [21] Logging must be done in accordance with otherwise applicable law, in particular data protection law and the rules concerning the protection of trade secrets.
>
> 22] The absence of logged information or failure to give the victim reasonable access to the information should trigger a rebuttable presumption that the condition of liability to be proven by the missing information is fulfilled.
>
> [23] If and to the extent that, as a result of the presumption under [22], the operator were obliged to compensate the damage, the operator should have a recourse claim against the producer who failed to equip the technology with logging facilities.
>
> [24] Where the damage is of a kind that safety rules were meant to avoid, failure to comply with such safety rules, including rules on cybersecurity, should lead to a reversal of the burden of proving
>
> (a) causation, and/or
>
> (b) fault, and/or
>
> (c) the existence of a defect.
>
> [26] Without prejudice to the reversal of the burden of proof proposed in [22] and [24](a), the burden of proving causation may be alleviated in light of the challenges of emerging digital technologies if a balancing of the following factors warrants doing so:
>
> (a) the likelihood that the technology at least contributed to the harm;

[27] Cf. Blanco-Justicia and Domingo-Ferrer (2019).

[28] Cf. Wojtczak and Księżak (2021).

> (b) the likelihood that the harm was caused either by the technology or by some other cause within the same sphere;
>
> (c) the risk of a known defect within the technology, even though its actual causal impact is not self-evident;
>
> (d) the degree of ex-post traceability and intelligibility of processes within the technology that may have contributed to the cause (informational asymmetry);
>
> (e) the degree of ex-post accessibility and comprehensibility of data collected and generated by the technology
>
> (f) the kind and degree of harm potentially and actually caused.
>
> [. . .] If there are multiple possible causes and it remains unclear what exactly triggered the harm (or which combination of potential causes at which percentage of probability), but if the likelihood of all possible causes combined, that are attributable to one party (e.g. the operator) exceeds a certain threshold (e.g. 50% or more), this may also contribute to placing the burden of producing evidence rebutting such first-hand impressions onto that party.

It is important to note that besides the technical measures, the proposal includes a number of strictly legal methods of managing the problem of unexplainability such as the reversal of the burden of the proof, lowering the standard of the proof, or both methods together. The necessity of such deviation is usually justified by the informational asymmetry between the disposer of the technology and the victim. Such methods have already been used in some systems of law in cases where problems have arisen with revealing and proving causal links, such as disputes on so-called medical malpractice. In Polish civil law doctrine, the following position is accepted:

> In "medical" proceedings, proving all stages of a causal link between an event indicated as causative and the damage may be extremely difficult or even impossible. Therefore, the case law admits the so-called *prima facie* evidence based on the construction of factual presumptions [. . .], which require the demonstration of high probability of the existence of the first and subsequent causative events, allowing to treat them as obvious. The causal link between the defendant's behaviour and the patient's death does not have to be established with certainty, a high degree of probability of the existence of such a link is sufficient, and in the case of a multitude of possible causes - an overwhelming probability of the causal link of the damage with one of these causes. (grounds of the verdict of the Supreme Court V CSKP 44/21 of 30 April 2021)

It is not the first time that the legal problems arising in the domain of AI seem to be analogous to those in medicine.

The measures postulated in the report from the Expert Group on Liability and New Technologies—New Technologies Formation, Liability for Artificial Intelligence and Other Emerging Digital Technologies (2019) are proposed in a similar form in Resolution 2020 and Proposal 2021, although they concern mainly high-risk AI-systems. One particularly important point is included in Article 62.1 of the Resolution 2020, which demands a form of self-denunciation by the provider of high-risk systems:

> Providers of high-risk AI systems placed on the Union market shall report any serious incident or any malfunctioning of those systems which constitutes a breach of obligations under Union law intended to protect fundamental rights to the market surveillance authorities of the Member States where that incident or breach occurred.

> Such notification shall be made immediately after the provider has established a causal link between the AI system and the incident or malfunctioning or the reasonable likelihood of such a link, and, in any event, not later than 15 days after the providers becomes aware of the serious incident or of the malfunctioning.

This provision not only requires the reporting of serious incidents or system malfunctions constituting a breach of obligations under EU law, which protects fundamental rights, but it also demands admitting the existence of a causal link between the AI system and the incident or malfunction, or at least the reasonable likelihood of such a link.[29] It is interesting whether such a system will work in practice, and how efficiently, but there is no doubt that information acquired in such a way could be a source of argument for eventual legal disputes and court cases. This is all the more likely given that, as it was mentioned at the beginning of this Chapter, Proposal DLDP 2022 and Proposal ALD 2022 both include the legal institution of disclosure of evidence applied in civil court procedures settling disputes over damages being the result of actions of AI. This institution is to be supported by rebuttable presumptions of product defectiveness and of the causal link between defectiveness of the product and the damage (Proposal DLDP 2022) and by the rebuttable presumptions of defendant's non-compliance with the duty of care and of the causal link between the fault of the defendant and the damage (Proposal ALD 2022).

Wojtczak and Księżak (2021) proposed various solutions to this problem. One would be the use of a system of governmental certificates or registers acknowledging the given AI a degree or class of safety, which would be interconnected with the procedural instruments of proving causality. Another may be changing the rules concerning the so-called social participation in legal proceedings. Since in many criminal law trials and some civil law trials in the common law system, the facts are decided by a jury, why not limit the power of professional judges and delegate the decision about the causal link to a jury, but a special one consisting of sworn experts? The traditional jury was originally introduced to give society surveillance over court decisions and to protect the citizens from the arbitrariness of state authorities, and since it consisted of laypeople, it also secured the equality of the law during court trials. Today, in such "unexplainable" cases as those involving the participation of AI, allowing non-professionals decide the causal link between the action of AI and the damage paradoxically would be an instrument of arbitrariness. If a mere mortal, even the judge, cannot understand the operation of an AI system, it is not possible for a just and legitimate decision to be made about it. To maintain equality between the plaintiff and the defendant, the facts about AI should be examined by such entities who are able to understand them. In such cases, the decision would not be an effect of several expert opinions competing one with each other in front of a judge

[29] Before reporting a provider must establish a causal link. So, when someone reports, simultaneously admits that there is a causal link. Usually, the other party would have to proof the causal link at the court.

(agonistic model) but the result of discussion and reflection made in a group of sovereign experts targeted to obtain the truth (deliberative model).

The task of recruiting sworn experts, to assure their objectivity and unbiased thinking, may be difficult, but not impossible. While they certainly should be human today, the role may be taken by an artificial entity endowed with AI in the future. This proposal may seem naïve and dispensable; after all, judges successfully decide on very difficult matters of facts in medical, pharmaceutical and technological areas, they can also decide on the facts where AI is involved; there are also many areas in which experts can also cope with uncertainty. However, the opacity of AI, the infamous 'black-box', presents a significantly greater problem than that of medical, pharmaceutical or technical issues, and is progressing. In addition, experts are more likely to identify a causal link than a judge. Furthermore, the impossibility of making a decision based on facts is nothing new (cf. the *non liquet* decisions in Roman law) and this issue has fuelled such ideas as those of broadly-understood proportional liability. With this in mind one can ask what is more just and fair: placing the decision about a causal link in the hands of a jury of experts, or in those of a judge. The former are more likely to identify the person responsible for the damage, and its extent. If this decision were left to a judge, the uncertainty may well result in the liability being divided between the probable tortfeasors.

11.5 Negligence: Standard of Conduct (Reasonable Care, Due Diligence and so on)—The Novelty for AI or Also for Humans?

The proposal included in the Resolution 2020 Annex B. point (18) relates to the issue of diligence which should be expected from an operator of an AI system. Such diligence, according to this prescription,

> should be commensurate with (i) the nature of the AI system: (ii) the legally-protected right potentially affected: (iii) the potential harm or damage the AI-system could cause: and (iv) the likelihood of such damage. Thereby, it should be taken into account that the operator might have limited knowledge of the algorithms and data used in the AI-system. It should be presumed that the operator has observed the due care that can reasonably be expected from him or her in selecting a suitable AI-system, if the operator has selected an AI-system which has been certified under a scheme similar to the voluntary certification scheme envisaged by the Commission. It should be presumed that the operator has observed the due care that can reasonably be expected from him or her during the operation of the AI-system, if the operator can prove that he or she actually and regularly monitored the AI-system during its operation and that he or she notified the manufacturer about potential irregularities during the operation. It should be presumed that the operator has observed the due care that can reasonably be expected from him or her as regards maintaining the operational reliability, if the operator installed all available updates provided by the producer of the AI-system. Since the level of sophistication of operators can vary depending on whether they are mere consumers or professionals, the duties of care should be adapted accordingly.

These demands seem to be analogous to those stated in PETL III.4.102.1 cited above; hence, the problems connected to due diligence appear simple and do not appear to diverge from the old ones.

However, it requires reflection, that compared to previous technologies and human craftmanship, the introduction of AI into common usage is leading to increased expectations regarding the outcome. Whether it is playing a game of Go, searching for information on the internet, driving a car or analysing X-rays, the quality of the result worked out by AI is in many aspects better than the one a human could deliver, such as safety, precision, quantity and productivity. Of course, this is not the case for every domain and every AI system. Even in those fields where the work is extremely advanced, the effects may still leave much to be desired; a good example would be the translation of a literary (artistic) text. However, our interest concerns the legal situation of those areas where emerging AI is better than humans, i.e. where it provides a higher level of relevant results, either now or in the nearest future. This phenomenon is worth examining because the use of AI entails higher expectations in its given field of application, and therefore, if one refers directly to private law constructions, to the expected level of care. And the consequences are indeed far-reaching.

Let's look at the example of the use of AI in medical diagnosis. Already today, some types of disease such as pneumonia, breast cancer and skin cancer, can be diagnosed by AI more effectively than by experienced doctors.[30] Progress in this area is very fast, and it can be expected that the number of cases where AI will surpass human abilities, sometimes significantly, will grow. This raises a question that can be formulated in two ways: what is the responsibility of the doctor, or healthcare facility, to use such an available system, or not,[31] and what is the standard expected from the doctor that affects his or her responsibility?

If harm occurs, and it was due to the fact that a hospital did not have a particular robot (or AI) due to a lack of money or resources, such as robots, then the answer would be the same as in the case of not having an expensive medicine or piece of equipment.[32] Solutions to such situations may vary between given legal systems, and may also depend on factors such as the patient's health insurance coverage. If, however, there was a (financial, factual and legal) possibility to use specific AI systems, but the health care institution, or the individual doctor, refused to do so or deliberately caused the patient to refrain from such a possibility, and this had a negative impact on the patient's health, such conduct should already be considered a culpable error in diagnosis or treatment, for which the institution or the doctor would

[30] Stanicki et al. (2021).

[31] O'Sullivan et al. (2019), p. 9.

[32] Although there is certain difference: while it cannot be argued that medicine or equipment (e.g. respirator or dialyzer) could be substituted by a human doctor, it can be rationally said that a human doctor could substitute an AI or robot in cases where, historically, it was the AI or robots which substituted human doctors. So, the argument may sound like this: "The hospital is not liable for damage caused by not having the surgical robot, because it had the best surgeon among its staff and he did what he could."

be held liable; depending on the situation, both tort and contractual liability may arise. In other words, once the use of certain AI tools is permitted, especially those related to diagnostics, assuming their price is not high and their capabilities exceed human ones, their use will become the standard, and the lack of such technological support will, in principle, imply malpractice. In this way, the issue of the standard expected of the physician will become obvious: the physician will have an obligation to use tools that provide better diagnostic or therapeutic effectiveness.

Both medical robots and AIs operating in the field of Medicine can be considered medical devices as they are understood by the law.[33] Such qualification necessitates compliance with very detailed requirements relating to the design, certification, labelling and registration of these devices, as well as monitoring their performance. The possibility of using AI and robots in Medicine requires official approval, as specified in Regulation (EU) 2017/75 on medical devices.[34] The conformity assessment and certification process for such devices is complex and is more akin to testing a new drug than a new tool. In the case of AI, it is intrinsically necessary to carry out tests to determine the precision of the machine, the probability of error, to detect situations where faulty decision-making may occur, and to determine the level of protection against hacking attacks. Careful and authoritative results can be obtained after experiments and clinical studies and IT tests. The certification of an autonomous machine as a medical device must guarantee not only the reliability of the machine in the given field of application, but also a high level of "competence" and "skill", and the statistical accuracy of the decisions made. This must be combined with a guarantee that the procedure for creating the AI eliminates any risk of bias or incomplete data, and that ensures legal and ethical actions at the operation stage.

Depending on the circumstances and application, the requirements for the certification of a medical AI may vary. However, it seems highly reasonable to expect

[33] According to Article 2 point 1 of Regulation (EU) 2017/75 on medical devices which is applied from 26.05.2020 'medical device' means any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes: "— diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease, — diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability, — investigation, replacement or modification of the anatomy or of a physiological or pathological process or state, — providing information by means of *in vitro* examination of specimens derived from the human body, including organ, blood and tissue donations, and which does not achieve its principal intended action by pharmacological, immunological or metabolic means, in or on the human body, but which may be assisted in its function by such means."

The following products shall also be deemed to be medical devices: "— devices for the control or support of conception; — products specifically intended for the cleaning, disinfection or sterilisation of devices as referred to in Article 1(4) and of those referred to in the first paragraph of this point."

[34] Regulation (EU) 2017/75 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223 and repealing Council Directives 90/385/EEC and 93/42/EEC, L 117/1, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017R0745&from=PL, last access on the 4th of August 2022.

that, as a general rule, AI should provide a standard no lower than that provided by a physician. Exceptions may apply, for example, to special types of applications, e.g. on a battlefield, where access to a human doctor will be impossible or difficult. In practice, due to psychological barriers, it can be expected that the only AIs allowed to be used are those that will unequivocally outperform even the best human professional. In a situation where the level of safety of AI will be by default higher than the level provided by the doctor, which will be the condition of certification, the expectation of the level of safety (or faultlessness) of a specific procedure will grow. Consequently, this will lead to pressure to eliminate the human factor where such a move will improve efficiency and safety. In other words, certification based on the assumption that the level of efficacy and safety of a particular AI is higher than that of a physician, actually immediately forces the use of that AI, if it is financially possible. For example, software that interprets X-rays at a level far higher than that of a human may be very cheap, and its release for use will simply force the immediate abandonment of human personnel, possibly reducing the medical agent to the ritual approval of AI results.

However, it is important to realise that human supervision over the results of AI work will in fact be based on irrational and extrinsic considerations: it is simply impossible to recognise the rationality of an activity in which a less effective, i.e. a more fallible, human controls an AI certified to have much higher effectiveness. Therefore, any such oversight should be focused on the process of the AI system entering the market, because afterwards it would be more logical that the AI should control the decisions made by the human. So, such *ex ante* control of AI systems, strongly determined by the state, should not be correlated with a more severe standard of liability on the part of its operator, as proposed by the Resolution 2020 and the Proposal 2021 (this was explained above).

It must be noted then, that making certain assumptions about the level of required diligence on the part of the AI results in a possible shift upwards in the level of required diligence on the part of humans. Therefore, should we strive to unify the level of diligence required of a human doctor and AI or should we acknowledge that from the doctor, we expect the highest human diligence; this assumes, among other things, that a human cannot react immediately, cannot manipulate big data and simultaneously access multiple variables, and that we can expect superhuman diligence from the AI. Even if, at the beginning, some doubts based on psychological reasons may arise, even if the fundamental rights impose today the rules, like the one in Article 22 section 1 GDPR,[35] on the grounds of the civil law an almost rhetorical question should be put: who would want to be treated by a doctor who is allowed by law to have a lower standard of service than that available? The consequences are obvious: the human doctor will not be able to meet the standard set by AI, and therefore will have to withdraw from the field occupied by the AI, if only for civil

[35] Article 22 section 1 GDPR: The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

liability reasons. The AI will simply do the job better. While its activity will be first restricted to certain areas, and under human supervision, its range will inevitably expand and ultimately, it will be the doctor who has to submit to the AI. People (patients and doctors) will notice that the number of errors made at certain stages of treatment decreases as human participation decreases,[36] and thus the introduction of AI will result in a fundamental paradigm shift in Medicine.[37]

As noted above, for psychological and social reasons, the law will require human supervision of AI operation in Medicine, even if it becomes just a meaningless ritual. The need to ensure that AI remains under human control and that the final decision rests with the human being is considered one of the cornerstones of the construction of rules relating to AI.[38] With the development of AI and the growing gap between its capabilities and those of humans, however, this postulate will become utopian, and so-called "human control" or "human oversight" will be reduced to ensuring that the AI operating procedures are followed correctly by the human operator rather than any actual monitoring of the AI itself. Even though the humans who will supervise the AI may believe that a certain outcome (e.g. a decision regarding treatment, diagnosis, etc.) is wrong, this cannot lead to the conclusion that it is the AI system that is wrong: after all, it will be more likely that the humans are wrong. This conflict, of course, will be complicated by the problems posed by overcoming the *black box* or *explainability* problem to justify the AI's decision, and the real possibilities of verifying such a decision.

It may be possible to generalise the above remarks on the issue of human diligence into the operation of AI in other areas. The introduction of AI will result in the standard (quality) of the expected decision (choice) made in a given circumstance rising to the point where humans will not be able to meet it. As a consequence, the following questions arise:

1. Can fault can be attributed to a person for the mere fact of not using AI if such technology is available?
2. Should the correct use of AI be regarded as a circumstance absolving guilt, if such use is permissible and possible in the circumstances?
3. Is it possible to tolerate a situation where AI and humans act simultaneously with different standards of care required?

Re. 1. The question of the consequences of non-compliance with AI decisions is, in our opinion, key to understanding the significance of the legal changes associated with the introduction of AI. Its essence is that unlike previous tools and inventions, AI no longer merely supports the execution of human will, but transfers the very decision-making centre from human to machine. This is unlike any change we have

[36] The same will happen in many other areas, such as road traffic. At the beginning the autonomous vehicles will operate under additional control demanded by the law, but over time, the possibility of human driving will be restricted because it will obviously increase the danger of accident.

[37] Hoeren and Niehoff (2018), pp. 308 and further.

[38] Hoeren and Niehoff (2018), pp. 308 and further; Schoenberger (2019), pp. 171–203.

previously faced following a new invention or, more broadly, scientific discovery. Of course, the invention of the axe, the telephone or the hoover has changed the circumstances of those using them or who are involved in providing certain services. While ever-improving diagnostic and surgical tools are raising the expected standard of treatment, it remains the doctor who decides whether and how to use them. Similarly, a lumberjack who once had to use only an axe and a saw now has an electric saw and even harvesters at his disposal, but ultimately it is the lumberjack who decides which tree to cut down and how. AI is, however, changing this paradigm: in Medicine, it will determine how to treat, and in forest management, it will no doubt indicate which trees should be cut down and how it should be done. This raises the question of whether humans can refuse to 'submit' to decisions made by an AI, and to what extent, or whether this step inevitably involves them accepting the risk of responsibility.

Evidence that such fears are present in reflections on AI can be seen in a paragraph included in the European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088(INI)). In point 77, the European Parliament:[39]

> Stresses, however, that the existing system for the approval of medical devices may not be adequate for AI technologies; calls on the Commission to closely monitor progress on these technologies and to propose changes to the regulatory framework if necessary in order to establish the framework for determining the respective liability of the user (doctor/professional), the producer of the technological solution, and the healthcare facility offering the treatment; points out that legal liability for damage is a central issue in the health sector where the use of AI is concerned; stresses the need therefore to ensure that *users will not be led invariably to back the diagnostic solution or treatment suggested by a technological instrument for fear of being sued for damages* if, on the basis of their informed professional judgement, they were to reach conclusions that diverged even in part [. . .]

The danger indicated here has been described very narrowly; it concerns only liability related to the application of AI in Medicine. However, the idea behind the formulation of point 77 has further consequences, and the reservation expressed in it may be generalised. The European Parliament notes that a human being, in this case a doctor, may be put in a situation where he or she will have to take a decision suggested by an AI, as refusal to do so will entail liability for damages; in this case—the doctor will be forced to apply a specific diagnostic solution or treatment.

Similar situation is examined by Hacker et al. (2020), pp. 423–424 and they say that:

> [. . .] even models superior to human judgment on average will generate some false negative and false positive recommendations. Hence, the use of the model should always only be part of a more comprehensive assessment, which includes and draws on medical experience [. . .] Doctors, or other professional agents, must not be reduced to mere executors of ML judgments. If there is sufficient, professionally grounded reason to believe the model is wrong in a particular case, its decision must be overridden. In this case, such a departure

[39] https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html, access on the 4th of August 2022.

from the model must not trigger liability – irrespective of whether the model was in fact wrong or right in retrospect.

This dilemma may not only concern issues related to the use of AI in Medicine, nor only the question of civil (or even criminal) liability. As one can guess, the European Parliament expects appropriate solutions that would ensure the doctor has freedom of choice in his duties or, in other words, to not be held responsible when he does not use an AI or chooses to ignore its advice. Let us leave aside here the problem of whether this is possible at all. It can only be noted in passing that it seems rather utopian, because it would mean that medically erroneous behaviour, despite the existence of tools for correct behaviour, would be deprived of sanctions.

Going back to the question posed earlier: who, apart from a group of some anti-AI-ers, would want to use the services of such a doctor? However, the problem is a more significant one: after all, this is only a special case illustrating a wider, basic issue: various AI systems will make decisions by which humans will be bound in one way or another. Even if theoretically, there is no such binding, as in the case of the doctor mentioned by the European Parliament, it will be a considerable personal risk for the user to contradict the decision of an AI. Of course, such a decision would also sacrifice the greater speed and convenience associated with automation.

However, should this psychological desire to make one's own decisions be supported by immunity from disciplinary, civil, criminal or political liability? If legal norms, including those of civil law or even a constitutional norm, still presuppose a decisive role for the human being, then this must presumably also entail human responsibility; but would anyone really have to bear that responsibility? Taking the example of autonomous financial control systems, from accounting programs to those analysing paths for optimal financing of health care; in such cases, is it reasonable to impose on an official "responsible" for a given area of administration the responsibility for the decisions made by such systems, considering that his role in the process is purely symbolic ?

It can be seen that this issue has far-reaching consequences: if AI decisions are accurate and correct, but some accepted fundamental norm guarantees humans a role in decision-making in any case, this guarantee must extend to allowing completely irrational human decisions when they are made in opposition to AI, and even protecting them. It is hard to imagine such a norm this way being applied in practice.

A broader reference to this problem can be found in Article 14 of Proposal 2021:

Article 14
Human oversight

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.
2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably

foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.

3. Human oversight shall be ensured through either one or all of the following measures:
   (a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;
   (b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.
4. The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances:
   (a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible;
   (b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias'), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;
   (c) be able to correctly interpret the high-risk AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available;
   (d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system;
   (e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure.
5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons.

The above provision commands the preparation of a range of measures for working with high-risk AI-systems, i.e. systems which would be strongly controlled according to the law and by the statal institutions; these would enable individuals in any concrete situation to decide not to use the high-risk AI-system or otherwise disregard, override or reverse its output, to be able to intervene in its operation or to interrupt the system through a "stop" button or a similar procedure. The consequences of such regulation may be easily understood when one imagines a cliché case: an AI system predicts the eruption of a volcano, its sends an evacuation signal to a nearby town but this is stopped by the mayor, who sees no signs of eruption and is afraid of losing tourists. A real example can be seen in Poland, when the government increased the air purity threshold for declaring a smog alert to four

times the objective norms applied in other EU countries to avoid frequent smog alerts tarnishing the image of the government.[40] Of course, if an AI system were working autonomously in such situations, it would proceed according to objective data, ignoring *ad hoc* private reasons.

There is also another side to this problem. If we consider that decision-making often takes place not when the interests of the decision-maker are being realised (e.g. a doctor or a social service worker buys a lunch for himself), but when the interests or rights of others are to be realised (e.g. the doctor decides about the treatment of a patient or the social service worker assists a domestic violence victim), the question arises whether these stakeholders (e.g. the patient, the woman) have the right to limit the autonomy of the will of the human decision-maker and to submit to the choice resulting from the expert opinion of an AI.[41] Should there not then be a constitutional right of the patient to be treated, if he chooses, strictly according to the indications of the AI? Should not a suspect in a criminal trial have a constitutional right to be declared innocent if the opinion of some duly authorised or certified AI system shows that he did not commit the act he is accused of, or that the evidence for committing the act is unreliable? Should there not be some kind of constitutional political right of the community of citizens that, in certain spheres, the decision of their representatives, parliamentarians or the government, should not be different from one resulting from the expert opinion of an AI?

For example, an AI could be used in the sphere of electoral law and the organisation of constituencies to ensure optimum representation of different groups of citizens,[42] or in the sphere of settling detailed criteria for tax burdens and social transfers, so that they best protect the values adopted in the constitution. It could also be used to disseminate data of importance to the community, e.g. on illnesses and mortality during pandemics, in the sphere of environmental protection by preventing the emergence of investments that are particularly harmful to the environment, or even in the area of health protection when distributing limited medical resources or appropriate medical facilities.

It may be said that such constitutional rights or guarantees would help to exclude a very dangerous and growing phenomenon that public authorities govern societies based on lies and misinformation. F. von Schirach has proposed the addition of six new fundamental rights to the Charter of Fundamental Rights of the European Union. Two of these directly concern AI (Article 2—Digital self-determination: Everyone has the right to digital self-determination. Excessive profiling or the manipulation of people is forbidden; Article 3—Artificial intelligence: Everyone

[40] https://airly.org/pl/alert-smogowy-kiedy-i-przy-jakich-wartosciach-oglaszany-jest-alarm-z-powodu-smogu/, last access on the 4th of August 2022.

[41] The problem is similar to that relating to so-called conscience clauses, when the view of one person, e.g., a doctor or pharmacist, may prevail over the claim of another person to receive a certain service.

[42] Of course, this problem is connected to political decisions and the concept of democracy accepted in a given time and a given place, but it does not mean that it should not be discussed or at least noticed.

has the right to know that any algorithms imposed on them are transparent, verifiable and fair. Major decisions must be taken by a human being.) while another, Article 4—Truth, has significant implications on AI development "Everyone has the right to trust that statements made by the holders of public office are true".[43] While the latter is intended to prevent the use of AI to disseminate lies by the holders of public office, it also ensures access by the individual to information collected with the help of AI.

In our opinion, although conclusions cannot be categorically formulated, the introduction of AI in a given area permanently changes the standard of care and legitimate expectations about the quality of any decisions that are made. Consequently, even if current norms declaratively assume the supremacy of the human will and its autonomy over the machine, there will be a *de facto* transfer of full autonomy (will) to machines in every subsequent field. However, in some time this inconsistency between what is being declared and what is actually happening may be noticed and eliminated. In such cases, the mere fact of not using an AI in certain processes may become the basis of attributing fault.

Re. 2. By accepting that the appearance of a certified AI in a given domain of activity *de facto* necessitates its use (as in the case of medical devices), it is also necessary to exempt the user (operator) of the AI from fault if the AI system malfunctions, provided it was properly certified and subsequently used. The user (operator) of such AI has exercised due diligence by entrusting the decision to the AI: nothing more is required of him. It is not at all certain that due diligence consisting of monitoring the work ("decision") of the AI is required. After all, we have established above that such demands are irrational and result from fear rather than logical analysis. If an AI playing chess or go indicates a certain move in a given position, which seems wrong to most of the world's best players, what does it actually do? If an AI analysing pictures of lungs claims that a patient is ill, while an experienced radiologist sees nothing of the sort in the picture, what does it conclude? Of course, it may happen that the human is right and the machine is wrong, this possibility of error is inherent in every system, but such errors are considerably less likely in the case of a certified AI than in the case of a human.

Therefore, the general possibility of error precludes the formulation of a rule based on the assumption of controlling the machine. Indeed, such errors should be detected at the design and certification stage, with the user's responsibility being limited to diligent care of the AI, i.e. following the manufacturer's instructions, such as updating software and undergoing mandatory testing. If these general duties are fulfilled, there is no longer any room for seeking fault in neglecting to check a particular decision made by the AI. In this case, a certain analogy can be drawn with having responsibility for a damage caused by a raised (kept) animal.

[43] https://www.jeder-mensch.eu/informationen/?lang=en, last access on the 4th of August 2022.

Negligence, on the other hand, can be attributed to negligent examination of the certification itself.[44] As a rule, the end-user will not be able to verify the operation of the AI, so he has to trust in the AI he uses. His liability may therefore only extend to the classic *culpa in eligendo*, i.e. fault in the choice of AI. However, careful selection of the AI, verification of its source and observance of the registration data should be sufficient to exclude liability.

Re. 3. The standard of due diligence may vary depending on the field in which the AI system operates. In some areas, such as medical activity, there may be only one standard, which would only be achievable by machines. However, in others, such as driving a car, different standards may exist for the AI and for humans (at least in the transitional period). In this latter case, autonomous vehicles will probably be subject to a higher expected level of care or caution than human drivers; despite the prevailing risk principle and mandatory insurance, this is of some importance. Under the same circumstances, an accident may be judged differently if it is caused by a human driver or a machine. It is to be expected that better and safer autonomous systems will gradually displace human drivers, resulting eventually in a complete ban on human driving; this replacement may well be hastened by providing appropriate economic incentives (e.g., from insurance premiums) for voluntary cessation. Doubtlessly, similar processes will also be seen in other areas.

The above reflections demonstrate that it is inevitable that the concepts of diligence and negligence will undergo change in response to the development of AI. Furthermore, these changes will be long lasting, at least until some new balance is reached between the activities dominated by advanced technological processes and those remaining in the human sphere.

11.6 Culpability of AI

11.6.1 Legal Culpability for AI: Why Is It Needed?

For the purpose of clarity, in this chapter, *culpability* is understood in a broad sense as a notion embracing both criminal and civil concepts of blame, and both intention and negligence in the field of civil law.

All European acts and studies on civil liability for damage caused by AI, and almost all postulates of doctrine concerning this matter, strongly insist that strict liability should be borne by the generally-understood holder of the AI, such as its

[44] The rules proposed by EU bodies insist on the rule of risk-management and give many criteria which should be applied to assess the risk of harm posed by AI. The instance may be here Article 7 of Proposal 2021, which establishes the directives for the Commission regarding the method of updating the list in Annex III with the addition of high-risk AI systems, and Article 9 which details a risk management system.

producer, owner, user, operator, provider or developer.[45] Such a model of strict liability, i.e. by holders, seems to better protect the interests of vulnerable subjects of the law, such as consumers. Furthermore, it is easier to apply and appears to be sufficient for today's needs. However, in the near future, it may be possible that:

1. strict liability would appear unfair, for example, when the given AI is being used in the public interest (aside from the profit it is accruing for its holder), because it diminishes the natural risk; in such a case, strict liability could be a negative stimulus for AI development;
2. a damage resulting from the action of one AI is placed in another AI.

Moreover, for some types of AI, their participation in the market and social life may be so efficient and autonomous that they would be endowed with some range of legal subjectivity in civil law; this status would be analogous to the subjectivity of juridical persons, for example. Indeed, the present monograph also postulates that some types of AI should be endowed with some range of subjectivity (see the chapter about legal subjectivity).

If AI were to have some scope of legal subjectivity, it could be responsible for its own behaviour, especially since the postulate of providing an AI legal subjectivity is often connected with allowing it to own property; this property may be used as a source of compensation. In such a situation, making some humans or other entities responsible for the action of such an AI, for example by imposing on them to strict liability, may seem an unnecessary injustice. However, in such cases, there needs to be a correct way of making AI legally responsible.

While elaborating the former situation, it should be kept in mind that the origins of the strict liability regime are associated with the problems which became apparent in the early industrial era, when the introduction of new machines and natural forces into common use increased the level of risk in social life. As this risk influenced both the users of the machines and natural forces, and others who did not benefit from their use, it was generally accepted that the users, who benefit from the use of the

[45] Among others: Directorate-General for Internal Policies Policy Department for Citizen's Rights and Constitutional Rights and Constitutional Affairs, Artificial Intelligence and Civil Liability. Study requested by the JURI Committee, PE 621.926—July 2020, Resolution 2017, Commission Staff Working Document: Liability for emerging digital technologies accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe, Brussels, 25.4.2018, COM (2018) 237 final, Report from the Expert Group on Liability and New Technologies—New Technologies Formation, Liability for Artificial Intelligence and Other Emerging Digital Technologies, European Union 2019, Report from the Commission to the European Parliament, the Council and the European and Social Committee: Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and Robotics, Brussels, 19.2.2020, COM (2020) 64 final, Resolution 2020, Hamon et al. (2020), EUR 30040.

machines or forces at the cost of somebody else's risk, should pay for it. This assumption has been unreflectively transferred into the debate about AI.[46]

Meanwhile, today we are beset by increasingly complicated situations. Some technologies dating from the industrial era have become so widespread and needed that even if they provide a direct benefit to a concrete person, they also benefit the whole of society and are part of a new standard of living. For example, electrical energy is essential for functioning of society, and its lack is treated on a par with natural catastrophes; in addition, cars are indispensable for providing immediate help in the case of a medical emergency or non-stop supplies of food and other goods.

The car superseded the horse-drawn carriage as the primary form of transportation around a century ago. Incidentally, the widespread use of horse-based transport was also associated with a greater risk than human transport had been. However, unlike older times, the development of the new and emerging technologies of the digital era is not so much intended to increase speed or power, but to improve accuracy, inerrancy, perfection and overall safety, as well as allow unattended operation. A good example is that of the autonomous car, the development of which does not lie in being faster than the previously-produced cars but in being unattended and safer. Another one is that of medical robots, which are not designed to increase the number of medical interventions per day (though of course it may be additional effect of introducing tireless technology) but to make them more accurate and safer.

Hence, the technological development of today diminishes many risks, especially those of physical damage (the issue of non-physical damages is more complicated); as such, the fruits of this development not only serve the interest of the individuals directly profiting from their use, but they also serve the public interest. And here lies the rub: if a patient prefers being operated on by a medical robot than a human surgeon, because the robot is more accurate and precise, why should the hospital be punished with the more severe regime of liability (strict liability) for using the robot than a human surgeon (fault liability)? If streets are safer, both in a statistical and individual perspectives, when only autonomous cars are travelling on them, why tighten the regime of liability for damages resulting from transport accidents, e.g., by diminishing the range of exemption circumstances? Such doubt is eminently justifiable; serious predictions have been made that autonomous cars will naturally displace regular ones, or human drivers will be limited or completely forbidden because of safety reasons. These examples show that the justifications for more severe standards of responsibility which were used so far are today not enough.

It is more probable that, as emerging technologies are used on increasingly greater scales, an AI may be damaged by another AI; in most cases, this would, legally speaking, result in economic loss for the AI's owner. This would be

[46] Resolution 2020, Annex, B. (8): [. . .] whoever creates, maintains, controls or interferes with the AI-system, should be accountable for the harm or damage that the activity, device or process causes. This follows from general and widely accepted liability concepts of justice, according to which the person who creates or maintains a risk for the public is liable if that risk causes harm or damage, and thus should *ex-ante* minimise or *ex-post* compensate that risk.

particularly true in the field of labour (in a general sense), where the human participation is predicted to decrease by a significant degree. Assume that one smart robot making a delivery destroys another smart robot which were to receive this delivery: it would be unjustified to apply a strict liability regime to this incident, as both robots were acting in the same temporal and technological conditions and thus were objects of the same legal and technological demands and standards.

11.6.2 Legal Culpability: The Concept Representing Physical Fact or Social Fact? Is It Possible for AI to be Culpable?

The prevalent view is that it is not possible to evaluate the acts performed by an AI against the criterion of legal culpability because in most legal systems, both in criminal and private law, awarding culpability requires the attribution of a certain attitude towards the deed or a certain negatively-perceived mental state to the subject.[47] Therefore, as contemporary AI is not conscious, it is not possible to attribute any mental state to it, and it cannot be classified as legally culpable or not (*Machina delinquere non potest*). From a logical perspective, it could be said that AI does not belong to the domain of "being legally culpable of". As a matter of fact, in western culture, the only objects which can be attributed with some mental state are living conscious creatures: a paradigmatic human being, i.e. a human of proper age and with a sane mind. However, this issue raises certain problems.

The natural consequence of the position described above is that to decide on legal culpability, it is necessary to have at least to some minimal knowledge of the mental state of the subject. However, does this assumption really work in practice?

Even if it is certain that a human has some natural capabilities of cognizing another human's mental state, it is still necessary to pretend that these capabilities play a role in the determination of culpability, i.e. (to be consistent) the process of determining the mental attitude of the perpetrator to their actions. If this were the case, any means by which these natural capabilities could be improved should be welcomed, as they would optimizing the legal process. However, the law treats any such improvements deriving from natural science with suspicion; for example, brain imaging is not used willingly by courts.[48] Why?

The strong dissonance between treating sanity and insanity (or the ability to be culpable and its lack) by the law and medical sciences is examined by Gazzaniga, a medical sciences representative and not a lawyer but just a person of common sense. On the one hand, there are concerns that the law is failing to fully exploit the knowledge and the tools provided by the contemporary neurosciences and cognitive

[47] Gless et al. (2016) and Dafni (2018).

[48] de-Wit et al. (2016).

sciences,[49] while on the other, there is resistance toward the use of neuroscientific methods such as brain imaging with MRF or FMR. In the case of the former, there are fears that mentally ill or mentally abnormal people, who despite not being able to decide about their own behaviour in full, may in fact be decided culpable by the law, particularly criminal law, which seems immoral and unhumanitarian.[50] The latter concerns are guided by the concern that such findings may reveal that nobody governs their own behaviour in full, and that this may allow dangerous criminals to avoid responsibility. For all these doubts, Gazzaniga has the perfect answer:

> It does not follow that a person with abnormal brain scan has abnormal behavior, nor is a person with abnormal brain automatically incapable of responsible behavior. Responsibility is not located in the brain. The brain has no area or network for responsibility. [. . .] the way to think about responsibility is that it is an interaction between people, a social contract. [. . .] responsibility is a contract between two people rather than a property of a brain, and determinism has no meaning in the context. Human nature remains constant, but out in the social world behavior can change.[51]

If the exact knowledge of physical facts going on in the brain is not necessary for deciding about the culpability it means that this concept does not represent such facts. If responsibility is a general social contract, culpability must also be a social concept. If the subject (1) understands what this contract demands, (2) knows how to fulfil its rules and, what more, (3) expects the other subjects to fulfil this contract in relation to him, this subject is capable of being legally culpable and responsible.

The above stance is supported by the fact that contemporary legal systems define legal subjects which do not generate any mental state and yet are regularly classified as legally culpable or not, these being juridical persons. Of course, this is still an area of debate. For example, in the domain of criminal law, there has been, and still is, a quite lively discussion between the jurists from different legal systems which can be labelled in the following way: *Societas delinquere potest?* or *Societas delinquere non potest?*. One of the arguments in this discussion is that a juridical person or collective entity cannot be a subject of criminal responsibility, because it cannot bear the blame.[52]

Kaplan (2016), p. 106, a widely-known expert on AI, answers this argument in the following way:

> In at least some of these cases, the corporation itself is considered to have moral agency because the institution is capable of understanding the consequences of its behavior and has a choice of actions (whether or not to commit the crime) [. . .] So can a computer program be a moral agent? It can because it meets the definition. There is no reason you can't write a program that knows what it is doing, knows it is illegal [. . .], and can make a choice as to

[49] Gazzaniga (2011), p. 222, cites R. Sapolsky, professor of neurology who said: "It is boggling that the legal system's gold standard for an insanity defense – M'Naughten – is based on 166-year-old science. Our growing knowledge about the brain makes notions of volition, culpability, and, ultimately the very premise of the criminal justice system, deeply suspect".

[50] See the cases in Edersheim et al. (2012), also Greely (2011) and Weiss and Watson (2015).

[51] Gazzaniga (2011), pp. 228, 252–253.

[52] Mueller (1957–1958), Jordaan (2003) and de Maglie (2005).

> what actions to take. There is nothing that requires a moral agent to "feel" anything about right or wrong – the requirement is simply that it knows the difference. For instance, to be held responsible for murder psychopaths need not feel that it's wrong to kill someone or experience remorse – indeed, they may disagree with the prohibition against murder – they simply have to know that society regards it as wrong. Without proper programming, machines are natural psychopaths, but they don't have to behave that way. It's entirely possible to program a machine to respect an ethical theory and apply it to facts, so it follows that machines can know right from wrong and make moral decisions.

At the end of this subchapter it is worth noting one more issue which addresses the doubts of those who are convinced that:

(a) the decisions of juridical persons are supported by human beings, who are their bodies or employees, and
(b) these are their (humans') blames which somehow (e.g. as an aggregate or as an emergent phenomenon) constitute the overall blame of a juridical person, and
(c) these are facts which make the ideas of the fault responsibility (criminal or civil) of the juridical persons more justified than the ideas of such responsibility of AI, where people are not present.

Consider the following simplified example. There is a company with a board of directors consisting of three persons, and their decisions require an absolute majority of validly cast votes. The board of directors was to decide to strengthen the construction of a building built by the company; however, the vote was not passed, and the building collapsed. The court acknowledges the company as being at fault for the damages which occurred; however, it was found that one board member voted for strengthening, another voted against and the third abstained. This shows that the role of human beings in the actions of the juridical person is not analogous to puppeteer and puppets. The inherent, and decisive, element of decisions made by juridical persons are not only the partial decisions and the intent of the members of the bodies, but also the rules governing the competence and the process of making decisions. The rules are elements of a categorically social character. As such, juridical persons should be seen as physical-social systems and their decisions, and their potential culpability as resulting from the actions of such systems.

Similarly, AI can also be seen as a physical-social system. Its elements are in some ways determined by human beings, i.e., the software and hardware environment and the intended purpose, which is inherently a social element, and in some ways not, especially when the AI is to some extent autonomous.

Moreover, in a world where legal culpability exists as a category, a human being is also a physical-social system. Neither human decisions nor culpability result from a single decisive physical centre, in our brain for instance, one biologically-determined decisive procedure or one incentive. Neither human decisions are determined only by physical or biological phenomena nor only by social phenomena. Both elements are necessary, and their proportion is impossible to determine.

Although the conclusion remains uncertain, a good hypothesis is that ability to be legally culpable is a property of physical-social systems.

11.6.3 Legal Culpability: The Unified Concept, Radial Concept, or Many Concepts? Is It Possible at All to Cognize the Culpability of AI?

It is a popular view of scientists and laymen that, even if an AI were conscious and could be regarded as "being legally culpable of" an act, its culpability cannot be conclusively determined because it is very difficult to learn how the decisions of AI are actually made. Of course, the limitations of AI transparency[53] represent only one potential source of legal problems. The present passage will discuss the problem of determining the culpability of the person responsible for the AI, or of the AI itself if it acts in a way which may be acknowledged as reprehensible from the perspective of the law.

Similar problems have been encountered for centuries. Since legal responsibility was connected to moral responsibility, culpability in the law was seen through the lens of the perspective of the judged subject toward the act itself. Even disregarding the difficult philosophical questions concerning the ability of living creatures to make autonomous and reasonable decisions and to be moral subjects,[54] it should be considered that to judge the mental attitude of a creature to its actions, it is necessary to have an insight into its mental state. If a creature is a human being, *prima facie*, there is no problem: a human being, although its comprehension is involved in different difficulties, is to some extent, mentally cognizable for another human being. However, although this very fact is indisputable, there are many opinions on such possibility and its sources:

> In the domain of cognitive science, theories of social cognition are dominated by two paradigms [. . .]. Theory [TT] hypothesizes that social cognition operates by the subject actively making inferences about other people in accordance with a folk psychology. Simulation theory [ST] comes in inferential and subpersonal forms. Inferential ST hypothesizes that social cognition operates on the model of the subject's own mind. Subpersonal or neural ST hypothesizes that mirror neuron system provides a rapid, unconscious mechanism for social cognition by activating regardless of the agent performing the action (whether it be oneself or another). Recently, an interaction theory (IT) theory of social cognition has been proposed as a challenge to these competitors. IT hypothesizes that the perception of other minds, intentions and beliefs is direct, noninferential, and interactive. IT objects to TT's and ST's representationalism and detachment from embodied interaction.[55]

The above hypotheses indicate that the natural sciences tend to confirm that one human can have an insight into another human's mental state. As a consequence, the legal practice of attributing the guilt or fault to a human seems to be valid from the scientific point of view. However, the difficulties connected with this position will be discussed below. Despite this, a problem arises when, because of the legal demand to

[53] There is huge literature on this issue, both technical and legal, see e.g. Blanco-Justicia and Domingo-Ferrer (2019), Brkan and Bonnet (2020) and Hamon et al. (2020).

[54] Cf. Shapiro (2006) and DeGrazia (2006).

[55] Neemeh (2018), p. 1.

decide on guilt, there is a need to have an insight into the mental states of a non-human creature. Indeed, it should be remembered that animals have previously been subjected to criminal prosecution and punished according to the law.[56] How is it possible to gain an insight into the mental attitude of an animal, and thus find it guilty? Similar examples can be seen in the period of transition from preindustrial to industrial societies when, because of the growing complexity of industrial processes:

> [. . .] the amplifying effect of the complexity was not resolved in a mere multiplication of the number of accidents involving damage. The transformation also regarded their intrinsic quality. Such new damaging facts more and more often were connected to technical and industrial data: the progressive consolidation of interaction between humans and machines in the process of industrial production made it hard to define the source from which the damaging facts emerged. Their matrix, in other words, become anonymous and the causal connection between a specific action and its outcomes more difficult to be identified and proved.[57]

Therefore, the following question arises: how can the mental state of a unit consisting of a machine and a man be determined, when this unit acts contrary to the law? Furthermore, how is it possible to determine the mental state of a juridical person, acting on the field of legal transactions, who commits a deed which would be acknowledged a criminal offence or a tort if it had been committed by a natural person?

This issue can be examined more easily if culpability is viewed not as a unitary concept, but as a radial one organized by convention with respect to a composite prototype. Looking at the linguistic *usus*, a rough thesis may be given that prototypic culpability consists of the following idealized cognitive models (ICM[58]): psyche, morality, responsibility, reprehension, individualism vs. collectivism, religion, originality vs. social relations, imperfection vs. perfection, mistake and law. These names given to the ICMs are of course only conventional labels proposed by us; the listing is not enumerative. The simplest support of this thesis is the number of lexemes and names included in the category of culpability and used for expressing different sets of ICMs:

- "blame" for reprehension, morality, imperfection ICM
- "guilt" for psyche, morality, reprehension, individualism, imperfection, criminal law, law ICM
- "collective guilt" and "collective responsibility" for psyche, morality, responsibility, reprehension, collectivism ICM
- "sin" for religion, reprehension, imperfection, originality ICM
- "fault" for mistake, responsibility, reprehension, social relations, law ICM
- "corporate fault" for mistake, reprehension, social relations, collectivism ICM

[56] Evans (2009).

[57] Monterossi (2020), p. 5.

[58] For the definition of radial category idealized cognitive models and composite prototype see Evans (2007), pp. 29, 104, 177–179.

- "negligence" for imperfection, mistake, responsibility, reprehension, social relations, civil law, law ICM
- "intentionality" for perfection, psyche, reprehension, responsibility, social relations, criminal law, law ICM.

If culpability is a radial category organized by convention with respect to the composite prototype, then it is quite natural that the members of this category are only connected with the family resemblance, and the category can change and develop when needed. Besides, in spite of traditional beliefs, even today, not all of the elements of the category of culpability are confined with "anthropomorphic moorings".[59] So it would be quite natural if a new, less prototypical, instance of the category of culpability appeared, for instance regarding AI.

11.6.4 Legal Culpability: An Autonomous or Relational Concept? How to Assess the Legal Culpability of an AI?

When the concept of liability is employed in contemporary civil law systems of different countries belonging to the Western legal culture, it is usually associated with a certain model. This model is quite precisely represented in DCFR, which distinguish two kinds of culpability and defines them in the following way:

> **VI. – 3.101: Intention**
> A person causes legally-relevant damage intentionally when that person causes such damage either:
>
> meaning to cause damage of the type caused; or
>
> by conduct which that person means to do, knowing that such damage, or damage of that type, will or will almost certainly be caused.

> **VI. – 3:102: Negligence**
> A person causes legally-relevant damage negligently when that person causes the damage by conduct which either:
>
> does not meet the particular standard of care provided by a statutory provision whose purpose is the protection of the person suffering the damage from that damage; or
>
> does not otherwise amount to such care as could be expected from a reasonably careful person in the circumstances of the case. (pp. 399–400)

[59] Gobert (1994), p. 409.

It can be also noticed, that when one causes the damage intentionally, one certainly acts in bad faith. However, if one causes the damage negligently, one acts in good faith, although the action is not correct from the perspective of certain standards. As stated in DCFR:

I. – 1:103: Good faith and fair dealing

(1) The expression "good faith and fair dealing" refers to a standard of conduct characterised by honesty, openness and consideration for the interests of the other party to the transaction or relationship in question.

(2) It is, in particular, contrary to good faith and fair dealing for a party to act inconsistently with that party's prior statements or conduct when the other party has reasonably relied on them to that other party's detriment.

These definitions clearly confirm Gazzaniga's insistence that legal culpability is not a matter of mental state but of social contract. According to this concept, those who betray another person's trust, or who say or do something other than declared, or do less than demanded by the law or less than declared are legally culpable and legally condemned. Culpability is not an autonomous concept about mental state, with some internal positive or negative value, but is a relational concept about the fairness of the behaviour of the entity in relation to some socially-accepted and legally-imposed criteria; it is about fairness in fulfilling the social contact.

But how to translate these ideas into the domain of emerging technologies?

We must first identify the most general conditions of the social contract regarding AI and the sources from which they derive their trust. Of course, a good question to start with concerns the identity of the party to the social contract. Fortunately, the answer is simple: the party of the social contract is the entity endowed with legal subjectivity which was legally burdened with the responsibility for a given kind of behaviour. This may be an AI, if it is endowed with legal subjectivity within the legally determined range, especially when it is realized that, as it was explained in the Chapter 2, legal subjectivity is always connected with participation in social life. If AI is capable of participating in social life, it is capable of being a party to the social contract.

We should next consider how an AI can be trusted by humans. It should not be disputable that our trust in AI technology is based on the following factual assumptions:

1. Every AI realizes a precisely-determined function which is socially accepted and declared, and not any other function.
2. Every AI is capable of realizing its function properly, with maximal effectiveness in given circumstances.
3. The above two conditions are controlled by state and the law.

The third condition is an issue discussed now by the public. For example, it has been proposed in Resolution 2017[60] that a special system of registration and certification should be introduced. The form of registration and certification remains generally undecided, except for the proposal included in the Proposal 2021 relating to high-risk AI, but it should be postulated that such a register should include a record of the function which is to be performed by a given AI (given as "intended purpose" in the Proposal 2021). This record should be accompanied by a description of the expected degree of effectiveness, either in the record or elsewhere: it should be openly recorded, accepted and accounted in the legal regulations that not all functions or purposes imposed on an AI may be performed with 100% certainty (e.g., the weather forecast).

It should also be noticed that even if a general (global, European or state) register is provided for the riskiest or the most influential AI systems, there should also be less significant (regional, local, industry) registers created for lower-risk systems, so that all AI systems are included in the registration/certification system. AI systems which are not registered or certified should be acknowledged as illegal and should be eliminated from the market. The issue of registration is described in Chapter 4.3.

In creating such legal institutions, we should acknowledge two legal presumptions:

1. Every AI realizes its declared function and not any other one.
2. Every AI is capable of realizing this function with maximal (declared) effectiveness in the given circumstances.

These legal presumptions should be, of course, rebuttable presumptions (*praesumptio iuris tantum*). It is indeed actually possible that the given AI:

1. Does not realize its function, but another one or
2. Realizes its function, but not effectively or improperly, resulting in it causing damage.

Rebutting the first presumption results in attributing intent to the AI; rebutting the second presumption results in attributing negligence to the AI. This idea is developed below.

Claiming that a given AI does not realize its function, as declared in the register or certificate, but some other function demands a rebuttal of presumption no. 1, thus proving unfairness and something similar to bad faith by the AI; in this case, the entities which use AI are considered to be misled, because they are convinced that AI realizes its declared function. This can be done if:

[60] The second of "General principles concerning the development of robotics and artificial intelligence for civil use" says that European Parliament: "2. Considers that a comprehensive Union system of registration of advanced robots should be introduced within the Union's internal market where relevant and necessary for specific categories of robots, and calls on the Commission to establish criteria for the classification of robots that would need to be registered; in this context, calls on the Commission to investigate whether it would be desirable for the registration system and the register to be managed by a designated EU Agency for Robotics and Artificial Intelligence".

(a) It is confirmed by the author of the AI (the author designed the declared function and can recognize the abnormality) or
(b) It is proved that AI was "hacked" or
(c) the circumstances indicate, with some high degree of probability determined by the law, that the AI does not realize its prescribed function but another one.

So, if one of the above premises is confirmed, it should be acknowledged that the AI demonstrates intent (it should be noticed that it is conceptual necessity that if someone acts in bad faith he acts intentionally).

However, if the given AI realizes its function, as declared in the register or certificate, but in an improper or ineffective manner,[61] it should be acknowledged negligent, unless:

(a) the improperness or ineffectiveness was caused by the circumstances which could not be influenced by the author of AI or AI itself, e.g. the accessible data set did not let the other result or
(b) the improperness or ineffectiveness was caused by the harmed person himself.

Taking into consideration the constructs of negligence present in contemporary private law and the fundamental assumptions of this work as the necessity of registration, one can say that the standard of care for each AI would be fixed individually by its appropriate record in the register or certificate. Of course, these standards can differ depending on the kind of the given AI and its model, but they would have lie within the general limits imposed by law. However, it should be insisted very strongly that these general, legally-imposed limits would not be not determined by civil law. They should be the part of the administrative law environment regulating the registration or certification of AI.

Although some of the proposed exculpation clauses may resemble the statutory exoneration clauses liberating a subject from the strict liability, the proposed model is a model of culpability. The difference is clear. As a matter of fact, damage is not a necessary circumstance to apply this model and attribute culpability; luckily for the AI or its holder, the damage could have been prevented by someone or avoided in some way, but the action could have been intentional or negligent.[62] Of course, the consequence in civil law would not be to pay compensation, but some other arrangements may take place, such as terminating the agreement between parties due to loss of trust or reputation.

[61] Proving the improper or ineffective realization of a function is a matter of civil procedure. It may depend on the opinion of experts or the report of an entity controlling registration or certification.
[62] It is analogical to Williams' and Nagel's moral luck. Williams (1982) and Nagel (1979).

References

Books and Articles

Anderson SL (2008) Asimov's "Three Laws of Robotics" and machine metaethics. AI Soc 22(4): 477–493. https://doi.org/10.1007/s00146-007-0094-5

Asimov I (1942) Roundaround. Astounding Science Fiction

Asimov I (1976) Bicentennial man. Ballantine Books, New York

Asimov I (1985) Robots and empire. Doubleday, New York

Barbrook R (2007) Imaginary futures: from thinking machines to the global village. Pluto Press, London

Beckers A, Teubner G (2021) The three liability regimes for artificial intelligence: algorithmic a cants, hybrids, crowds. Hart, Oxford

Blanco-Justicia A, Domingo-Ferrer J (2019) Machine learning explain ability through comprehensible decision trees. In: Holzinger A, Kieseberg P, Min Tjoa A, Weippl E (eds) Machine learning and knowledge extraction. Springer, Cham. ISBN 978-3-030-29726-8

Bostrom N (2014) Superintelligence: paths, dangers, Strategies. Oxford University Press, Oxford

Brkan M, Bonnet G (2020) Legal and technical feasibility of the GDPR's quest for explanation of algorithmic decisions: of black boxes, white boxes and fata morganas. Eur J Risk Regul. 11(18): II.2–II.3. ISSN 2190-8249

Clarke R (1994) Asimov's laws of robotics: implications for information technology. IEEE Comput 27(1):57–66

Dafni L (2018) Could AI agents be held criminally liable: artificial intelligence and the challenges for criminal law. South Carolina Law Rev 69:677. https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=4253&context=sclr, last access on the 4th of August 2022

De Maglie C (2005) Models of corporate criminal liability in comparative study. Washington Univ Global Stud Law Rev 4:547

DeGrazia D (2006) On the question of Personhood beyond Homo sapiens. In: Singer P (ed) The defense of animals. Second Wave. Blackwell Publishing, Malden, pp 40–53

de-Wit L, Alexander D, Ekroll V et al (2016) Is neuroimaging measuring information in the brain? Psychon Bull Rev 23:1415–1428. https://doi.org/10.3758/s13423-016-1002-0

Edersheim JG, Weintraub Brendel R, Price B (2012) Neuroimaging, diminished capacity and mitigation. In: Simpson JR (ed) Neuroimaging in forensic psychiatry. From clinic to the courtroom. Wiley and Sons

Evans V (2007) A glossary of cognitive linguistics. Edinburgh University Press, Edinburgh

Evans EP (2009) The criminal prosecution and capital punishment of animals. The Lawbook Exchange Ltd., Clark, NJ

Floridi, L, Cowls, J, Beltrametti, M, Chatila, R, Chazerand, P, Dignum, V, Luetge, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E (2018) "AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and Recommendations", forthcoming in Minds and Machines, December 2018. https://ssrn.com/abstract=3284141, last access on the 4th of August 2022

Foot P (1967) The problem of abortion and the Doctrine of the double effect. Oxford Rev 5

Gazzaniga M (2011) Who is in charge? Free will and the science of the Brain. HarperCollins Publishers, Pymble (Australia)

Gless S, Silverman E, Weigend T (2016) If Robot cause harm, who is to blame? Self-driving cars and criminal liability. New Crim Law Rev Int Interdiscip J 19(3):412–436

Gobert J (1994) Corporate criminality: four models of fault. Legal Stud 14(03). https://doi.org/10.1111/j.1748-121x.1994.tb0510.x

Greely HT (2011) Neuroscience and criminal responsibility: proving 'Can't Help Himself' as a narrow bar to criminal liability. In: Freeman M (ed) Law and neuroscience. Current Legal Issues 2010 vol. 13. Oxford University Press, Oxford

Hacker P, Krestel R, Grundmann S, Naumann F (2020) Explainable AI under contract and tort law: legal incentives and technical challenges. Artif Intell Law 228:415–439

Hage J (2017) Theoretical foundations for the responsibility of autonomous agents. Artif Intell Law 3(25)

Hamon R, Junklewitz H, Sanchez I (2020) Research Centre Technical Report. Robustness and explainability of artificial intelligence – from technical to policy solutions. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/57493. (online), JRC119336

Hoeren T, Niehoff M (2018) Artificial intelligence in medical diagnoses and the right to explanation. Eur Data Prot Law Rev 4:3. https://doi.org/10.21552/edpl/2018/3/9

Jordaan L (2003) New perspective on the criminal liability of corporate bodies. Acta Juridica 48

Kaplan J (2016) Artificial intelligence – what everyone needs to know. Oxford University Press, Oxford

Księżak P, Wojtczak S (2020) Prawa Aismova, czyli science fiction jako fundament nowego prawa cywilnego. Forum Prawnicze 4(60). https://doi.org/10.32082/fp.v0i4(60).378

Lucas J (1993) Responsibility. A Clarendon Press Publication, Oxford

McCauley L (2007) AI Armageddon and the Three Laws of Robotics. Ethics Inf Technol 9(2): 153–164. https://doi.org/10.1007/s10676-007-9138-2

Monterossi MW (2020) Liability for the fact of autonomous artificial intelligence agents. Things, agencies and legal actors Global Jurist 20190054, eISSN 1934-2640, https://doi.org/10.1515/gj-2019-0054

Mueller GOW (1957–1958) Mens Rea and the corporation: a study of the Model Penal Code position on corporate criminal liability. Univ Pittsburgh Law Rev 19:21–50

Murphy R, Woods D (2009) Beyond Asimov: the three laws of responsible robotics. IEEE Intell Syst 24(4):14–20. https://doi.org/10.1109/MIS.2009.69

Nagel T (1979) Moral luck. In: Nagel T (ed) Mortal questions. Cambridge University Press, Cambridge

Nathan MJ (2021) Black Boxes: how science turns ignorance into knowledge. Oxford University Press

Neemeh ZA (2018) Husserlian empathy and simulationism, memphis: organization of phenomenological organizations VI: Phenomenology and Practical Life 2018, https://www.memphis.edu/philosophy/opo2019/pdfs/neemeh-zach.pdf, last access on the 4th of August 2022

Nilsson N (2009) The quest for artificial intelligence: a history of ideas and achievements. Cambridge University Press, New York

O'Sullivan S, Nevejans N, Allen C, Blyth A, Leonard S, Pagallo U, Holzinger K, Holzinger A, Sajid MI, Ashrafian H (2019) Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. Int J Med Robot Comp Assisted Surg 15(1):1–12. https://doi.org/10.1002/rcs.1968

Schoenberger D (2019) Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. Int J Law Inf Technol 27:2

Searle JR (1984) Minds, brains and science. Harvard University Press, Cambridge

Shapiro P (2006) Moral agency in other animals. Theoret Med Bioeth 27(4):357–373

Skinner BF (1969) Contingencies of reinforcement: a theoretical analysis. Appleton-Century-Crofts, New York

Stanicki P, Nowakowska K, Piwoński M, Żak K, Niedobylski S, Zaremba B, Oszczędłowski P (2021) The role of artificial intelligence in cancer diagnostics - a review. J Educ Health Sport 11(9):113–122. https://doi.org/10.12775/JEHS.2021.11.09.016

Vinge V (2003) Technological Singularity. http://cmm.cenart.gob.mx/delanda/textos/tech_sing.pdf, last access on the 4th of August 2022

Weiss KJ, Watson C (eds) (2015) Psychiatric expert testimony. Emerging applications. Oxford University Press, Oxford
Williams B (1973) A critique of utilitarianism. In: Smart J, Williams B (eds) Utilitarianism for and against. Cambridge University Press, Cambridge
Williams B (1982) Moral Luck [in] Moral Luck. Philosophical Papers 1973-1980. Cambridge University Press, Cambridge
Wojtczak S, Księżak P (2021) Causation in civil law and the problems of transparency in AI. Eur Rev Priv Law 29(4):561–582

Documents

Proposal for a directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM (2022) 496 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496
Proposal for a directive of the European Parliament and of the Council on liability for defective products, COM (2022)495 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A495%3AFIN&qid=1664465004344
Commission Staff Working Document: Liability for emerging digital technologies accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe, Brussels, 25.4.2018, COM (2018) 237 final, https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A52018SC0137, last access on the 4th of August 2022
European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088 (INI)), P8_TA (2019) 0081. https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html, last access on the 4th of August 2022
Evas T (2020) Civil liability regime for artificial intelligence. European added value assessment. Study. European Parliamentary Research Service. September 2020. Brussels. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654178/EPRS_STU(2020)654178_EN.pdf, last access on the 4th of August 2022
Expert Group on Liability and New Technologies – New Technologies Formation. Liability for artificial intelligence and other emerging technologies. 2019. https://doi.org/10.2838/573689. https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2020/01-09/AI-report_EN.pdf, last access on the 4th of August 2022
Open Letter to the European Commission: Artificial Intelligence and Robotics, http://www.robotics-openletter.eu, last access on the 4th of August 2022
Regulation (EU) 2017/75 of the European Parliament and of the Council of % April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223 and repealing Council Directives 90/385/EEC and 93/42/EEC, L 117/1, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017R0745&from=PL, last access on the 4th of August 2022
Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee. Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. 19.2.2020. Brussels. https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX%3A52020DC0064, last access on the 4th of August 2022

Chapter 12
Conclusions

It is certainly a truism to say that the law has not kept pace with technological developments; and this old truth becomes particularly relevant in the field of artificial intelligence. The discussion on the need for regulation has gathered exceptional momentum in recent years, as it has become increasingly clear that the existing legal framework is simply inadequate in many areas. However, the current debate on AI is still targeted towards formulating proposals concerning the fundamentals: the ethics of AI design and operation, the risks involved, the protection of fundamental rights, the protection of general data and, more broadly, human dignity, autonomy, subjectivity and privacy. In our opinion, despite its necessity, such reflection is no longer sufficient. A technological revolution is taking place in front of our eyes that will have a significant impact on all areas of law. Artificial Intelligence must be incorporated in some way into civil law, with its distant roots in Roman law. In our view, this cannot be done through cosmetic changes, relating only to certain areas (e.g. civil liability or copyright law): not only are such changes insufficient, they do not create any coherent picture. The debate regarding the creativity of robots cannot, in our view, exist in isolation from its counterpart concerning the legal capacity of AI or their personal rights: the discussion on the subjectivity of agents cannot leave aside issues of property, among others. However, it is also not correct to incorporate Artificial Intelligence "by force" into known constructs. Such shortcuts are evident, for example, in the idea of giving AI subjectivity, which, one could argue, would solve all problems in one go, *deus ex machina*. Nor can we believe that AI is just another new tool (technology) that can be described without difficulty through already known constructs. We believe the matter is much more complicated, and the emergence of AI, an autonomous entity with superhuman capabilities, whose field of action is constantly expanding, will have far-reaching consequences for civil law transactions. The emergence of AI will be a meteorite hitting the legal system, and it should be treated as such; to this extent, there is a need for a comprehensive, coherent concept covering private law, and this book serves to present a possible approach to achieving this.

P. Księżak, S. Wojtczak, *Toward a Conceptual Network for the Private Law of Artificial Intelligence*, Issues in Privacy and Data Protection 51,
https://doi.org/10.1007/978-3-031-19447-4_12

The starting point of this approach must be to address the crux of the discussion concerning AI: whether it can be treated as a subject of law. The discussion relating to embodied AI (robots) repeatedly raises the subjectivity or rights of robots as a theoretical and legal construct. We believe that in the context of civil law, the issue should be viewed differently: assuming under civil law that AI is a subject is equivalent to saying that it has some rights, which in turn means that another entity has some obligations. In civil law relations, this should be referred to the social function understood to be performed by a particular entity. If AI can enter into social relations to some extent, this should be recognized by civil law. This does not mean, however, that AI is to have the same or even a similar position as natural or legal persons. Its legal capacity under private law must be related to the purposes it serves. These purposes, in turn, should follow directly from the specification of the AI itself. From this, we conclude that allowing AI to operate in the market as something more than a tool in the hands of humans, one that merely transmits their will, requires full state control, expressed in a certification and registration process. The purpose stated in the registration documents should be used to determine the legal capacity of the AI.

A further consequence is the placement of AI in specific subsystems of private law, i.e., regulations relating to property, contract law or tort law. This also makes it possible to understand that, to some narrow extent, AI may have its own non-property interests (personal goods), including, in line with the actual state of affairs, the right to authorship of a work. "Robot rights", however, are never a value in themselves; they are a function of the role machines play in human society. Therefore, an AI's own right to existence, integrity or replication should be firmly excluded. At the same time, we believe that AI, as an autonomous entity, cannot be treated as a mere tool in the hands of humans, while also not being treated as an 'ordinary' entity, a mere participant in the marketplace. Thanks to its punctual legal capacity, there is no contradiction in AI being both an object and a subject depending on the context: both an object of ownership and an owner, and both an object of somebody else's responsibility and a subject that can be responsible itself.

At the same time, autonomous action must be recognised in the context of private law as a relevant circumstance wherever the subjective elements of the trader are relevant. Since AI not only transmits the will of a user, but also shapes it, it is both possible and necessary to apply to AI *mutatis mutandis* the concepts such as consciousness, knowledge, will, good faith and guilt. However, these concepts must be understood in a specific way when interpreted in the context of AI: the basic orientation must always be the purpose of the system, as defined in the registration documents.

We hope that our proposal will serve as a starting point for further discussion on the integration of AI into private law. We would like to draw attention to the need for a holistic view, to create a legal framework that encompasses all the changes taking place before our eyes. This network of systemic relations in civil law cannot be ignored when creating detailed solutions. While our proposals look to the future, the time to start building the system is today, and such construction should begin from the foundations, which we hope this book has in some part laid.