Chapter 2

Regulating AI and Robotics: Ethical and Legal Challenges

Martin Ebers*

in: *Martin Ebers/Susana Navas Navarro* (eds.), Algorithms and Law, Cambridge, Cambridge University Press, 2019 (forthcoming).

Version: 17 April 2019

Abstract: Rapid progress in AI and robotics is challenging the traditional boundaries of law. Algorithms are widely employed to make decisions that have an increasingly far-reaching impact on individuals and society, potentially leading to manipulation, biases, censorship, social discrimination, violations of privacy and property rights, and more. This has sparked a global debate on how to regulate AI and robotics.

The purpose of this introductory chapter is twofold. First, it outlines some of the most urgent ethical and legal issues raised by the use of self-learning algorithms in Artificial Intelligence (AI) systems and (smart) robotics. Secondly, it provides an overview of several key initiatives at the international and European levels on forthcoming AI ethics and regulation. The overall aim of this chapter is *not* to find definitive answers. On the contrary, the policy debate would be better served by refraining from rash solutions. What is needed instead is a more precise inventory of the *concrete* ethical and legal challenges to strengthen the foundations for evidence-based governance in the future.

* This work has been supported by the Estonian Research Council grant no PRG124. All internet sources referred to were last accessed on April 17, 2019.

H. (International) Initiatives to Regulate AI and Robotics

I. Overview
II. European Union
    1. The European Parliament's Resolution of February 2017
    2. The European Economic and Social Committee's Opinion on AI of May 2017
    3. The European Commission's AI Strategy and The Work of the High-Level Expert Group on AI
    4. Next Steps
III. International Organizations
    1. Council of Europe
    2. OECD
    3. United Nations
IV. Industry Initiatives and Self-regulation at International Level

I. Governance of Algorithms: Regulatory Options

I. Should AI Systems and Robotics be Regulated by Ethics or Law?
II. General Regulation versus Sector-specific Regulation
III. Guiding Questions For Assessing the Need to Regulate
IV. Level of Regulation: Global, International, National or Regional?
V. Instruments for Modernizing the Current Legal Framework
VI. A Plea for an Innovation-friendly Regulation

J. Outlook

A. Scenario

I. The Use of Algorithms by Businesses and Governments

Algorithms permeate our lives in numerous ways, performing tasks that until recently could only be carried out by humans. Modern Artificial Intelligence (AI) technologies based on machine learning algorithms and big-data-powered systems can perform sophisticated tasks – such as driving cars, analyzing medical data, or evaluating and executing complex financial transactions – without active human control or supervision. Algorithms also play an important role in everyday decisions. They influence nearly every aspect of our lives:

- *Self-learning algorithms* determine the results of web searches, select the ads and news we read, and decide which purchase offers are made when we shop online.[1]
- *Dynamic pricing algorithms* automatically evaluate events on (online) markets so that traders can adjust their prices to the respective market conditions in milliseconds.[2]
- Software agents *optimize portfolios*, assess credit risks, and autonomously carry out the most favorable transactions in currency trading. On the financial markets, algorithmic trading (including high-frequency trading) generates more than 70 % of the trading volume. In the FinTech market, Robo-Advisors are used for investment advice, brokerage, and asset management.[3]
- Algorithms also play an increasing role in *making substantive decisions*. Many important decisions which were historically made by people are now either made by computers or at least prepared by them. We live in a "scored society" (*Citron/Pasquale*).[4] Companies from various industries collect, analyze, acquire, share, trade, and utilize data on billions of people in order to discern patterns, predict the likely behavior of people through scoring systems, and act accordingly. Some algorithmic scores have existential consequences for people: They decide to an increasing extent whether someone is invited for a job interview, approved for a credit card or loan, or qualified to take out an insurance policy.
- *Governmental institutions* have become increasingly dependent on algorithmic predictions. Tax offices have started using algorithms to predict abuse and fraud in tax returns and to allocate cases for human review.[5] Criminal law enforcement agencies use algorithms to detect, respond to, and predict crime (predictive policing).[6] In the US, algorithmic prognosis instruments are already being used by courts to calculate the likelihood of an accused person

---

[1] *Christl*, Corporate Surveillance in Everyday Life. How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions. A Report by Cracked Labs, June 2017, http://crackedlabs.org/en/corporate-surveillance.

[2] *Chen/Mislove/Wilson*, An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace, (2016) Proceedings of the 25th International Conference on World Wide Web 1339-1349, www.ccs.neu.edu/home/amislove/publications/Amazon-WWW.pdf.

[3] *BI Intelligence*, The Evolution of Robo-Advising: How automated investment products are disrupting and enhancing the wealth management industry, 2017; *Finance Innovation and Cappuis Holder & Co.*, Robo-Advisors: une nouvelle réalité dans la gestion d'actifs et de patrimoine, 2016*; OECD*, Robo-Advice for Pensions, 2017.

[4] *Citron/Pasquale*, The Scored Society: Due Process for Automated Predictions*,* (2014) 89 Washington Law Review 1.

[5] *DeBarr/Harwood*, Relational Mining for Compliance Risk, Presented at the Internal Revenue Service Research Conference, 2004, http://www.irs.gov/pub/irs-soi/04debarr.pdf [https://perma.cc/Y9F8-RWNK].

[6] *Barrett*, Reasonably Suspicious Algorithms: Predictive Policing at the United States Border, (2017) 41(3) N.Y.U. Review of Law & Social Change 327; *Ferguson*, Predictive Policing and Reasonable Suspicion, (2012) 62 Emory L.J. 259, 317; *Rich*, Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment, (2016) 164 University of Pennsylvania Law Review 871; *Saunders/Hunt/Hollwood*, Predictions Put Into Practice: A Quasi Experimental Evaluation of Chicago's Predictive Policing Pilot, (2016) 12 Journal of Experimental Criminology 347.

committing another crime on parole.[7] In China, the government plans to implement by 2020 a Social Credit System which is intended to standardize the assessment of citizens' and businesses' economic and social reputations.[8]

- Another important sector in which AI systems are used is the health sector: *Medical expert systems* based on self-learning algorithms evaluate the medical literature and personal data of patients, assisting physicians with their diagnosis and treatment, whether by reading medical images and records, detecting illnesses, predicting unknown patient risks, or selecting the right drug.[9]
- To an increasing extent, embodied AI systems also operate physically in the world. They have left the factories and come into our lives as intelligent robotic assistants, vacuum cleaners, drones, and automated cars. AI systems are also an essential component of developing the emerging "Internet of Things" (IoT)[10] – a network of physical devices which are embedded with electronics, software, sensors, and network connectivity that enable them to collect and exchange data.
- Last but not least, new devices make it possible to connect the human brain to computers. *Brain-computer interfaces (BCI)* enable information to be transmitted directly between the brain and a technical circuit. In this way, it is already possible for severely paralyzed persons to communicate with a computer solely through brain activity.[11] Researchers at Elon Musk's company Neuralink predict that machines will be controlled in the future solely by thoughts.[12] What's more, Facebook is researching a technology that sends thoughts directly to a computer in order to make it possible to "write" one hundred words per minute without any muscle activity.[13] Thus, the boundary between man and machine is becoming blurred. Human and machine are increasingly merging.

The technological change triggered by AI and smart robotics raises a number of unresolved ethical and legal questions, discussed in detail below (cf. C.-G.). Before addressing these issues more fully, it is important to take a closer look at the question of what we actually mean when we speak of "algorithms, AI and robots", whether common definitions are necessary from a legal point of view (II.),

---

[7] Such processes are used at least once during the course of criminal proceedings in almost every US state; *Barry-Jester/Casselman/Goldstein*, The New Science of Sentencing, The Marshall Project, April 4, 2015, https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing#.xXEp6R5rD. More than 60 predictive tools are available on the market, many of which are supplied by companies, including the widely-used COMPAS system from Northpointe.

[8] *Hvistendahl*, In China, a Three-Digit Score Could Dictate Your Place in Society, Wired, December 14, 2017, https://www.wired.com/story/age-of-social-credit/; *Botsman*, Big data meets Big Brother as China moves to rate its citizens, Wired UK, October 21, 2017, http://www.wired.co.uk/article/chinese-government-social-creditscore-privacy-invasion.

[9] *Abu-Nasser*, Medical Expert Systems Survey, (2017) 1 (7) International Journal of Engineering and Information Systems 218; *Gray*, 7 amazing ways artificial intelligence is used in healthcare, September 20, 2018, https://www.weforum.org/agenda/2018/09/7-amazing-ways-artificial-intelligence-is-used-in-healthcare.

[10] The combination of AI, advanced robots, additive manufacturing, and the Internet of Things will combine to usher in the Fourth Industrial Revolution; *World Economic Forum*, Impact of the Fourth Industrial Revolution on Supply Chains, October 2017, https://www.weforum.org/whitepapers/impact-of-the-fourth-industrial-revolution-on-supply-chains.

[11] *Blankertz*, The Berlin brain – computer interface: accurate performance from first-session in BCI-naïve subjects, (2008) 55 IEEE Transactions on Biomedical Engineering 2452, doc.ml.tu-berlin.de/bbci/publications/BlaLosKraDorCurMue08.pdf; *Nicolas-Alonso/Gomez-Gil*, Brain Computer Interfaces, (2012) 12 (2) Sensors 1211, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3304110/.

[12] https://www.theverge.com/2017/2/13/14597434/elon-musk-human-machine-symbiosis-self-driving-cars.

[13] https://techcrunch.com/2017/04/19/facebook-brain-interface/?guccounter=1.

and, more generally, how AI systems and advanced robotics differ fundamentally from earlier technologies, making it so difficult for legal systems to cope with them (B.).

II. Concepts and Definitions

1. Algorithms, AI and Robots: Do we need all-encompassing definitions?

Algorithms are by no means new. For decades, they have served as integral components of every computer program. Generally speaking, an algorithm can be understood as "sets of defined steps structured to process instructions/data to produce an output".[14] From this point of view, every software is composed of algorithms.

This definition is on the one hand too broad and on the other hand too narrow, since a purely technical understanding of algorithms as computer code does not go far enough in assessing their legal and social implications. As *Kitchin*[15] rightly points out, algorithms „cannot be divorced from the conditions under which they are developed and deployed". Rather, "algorithms need to be understood as relational, contingent, contextual in nature, framed within the wider context of their socio-technical assemblage".[16]

Popular definitions of AI are equally unrefined.[17] AI is a catch-all-term, referring to the broad branch of computer science that studies and designs intelligent machines.[18] The spectrum of applications using AI is already enormous to date, ranging from virtual assistants, automatic news aggregation, image and speech recognition, translation software, automated financial trading, and legal eDiscovery to self-driving cars and automated weapon systems.

---

[14] *Kitchin*, Thinking critically about and researching algorithms, (2017) 20(1) Information, Communication and Society 1-14. According to *Miyazaki*, the term "algorithm" emerged in Spain during the 12th century when scripts of the Arabian mathematician Muḥammad ibn Mūsā al-Khwārizmī were translated into Latin. These scripts describe "methods of addition, subtraction, multiplication and division with the Hindu-Arabic numeral system". Thereafter, "algorism" meant "the specific step-by-step method of performing written elementary arithmetic"; *Miyazaki*, Algorhythmics: Understanding micro-temporality in computational cultures, (2012) 2 Computational Culture, http://computationalculture.net/article/algorhythmics-understanding-micro-temporality-in-computational-cultures.

[15] *Kitchin*, (2017) 20(1) Information, Communication and Society 1; *Seaver*, Algorithms as culture: Some tactics for the ethnography of algorithmic systems, (2017) July-September Big Data & Society 1, suggested thinking of algorithms not "in" culture, but "as" culture: part of broad patterns of meaning and practice that can be engaged with empirically. *Dourish*, Algorithms and their others: Algorithmic culture in context, (2016) July-September Big Data & Society 1, at p. 3, notes that "the limits of the term algorithm are determined by social engagements rather than by technological or material constraints".

[16] Cf. also *infra*, B.IV, with reference to three dimensions that can be found in every ADM system, i.e. the process level, the model level, and the classification level.

[17] The High Level Expert Group on AI (AI HLEG), set up by the EU Commission, proposes the following updated definition: "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions"; *AI HLEG*, A Definition of AI: Main Capabilities and Disciplines, Brussels, April 9, 2019, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341.

[18] *McCarthy*, What is Artificial Intelligence?, 2007, http://www-formal.stanford.edu/jmc/whatisai/. Russell and Norvig summarize eight definitions of AI differentiated by how they reflect expectations of human thinking and behavior or (machine) rational thinking and behavior; *Russel/Norvig*, Artificial Intelligence: A Modern Approach, 3rd ed., 2011.

From a legal standpoint, this lack of definitional clarity is sometimes regarded as problematic. Scholars emphasize that any regulatory regime must define what exactly it is that the regime regulates, and that we must therefore find a common definition for the term "artificial intelligence".[19] Others believe that an all-encompassing definition is not necessary at all, at least for the purposes of legal research and regulation.[20] After all, AI systems pose very different problems depending on who uses them, where, and for what purpose. For example, an autonomous weapon system can hardly be compared to a spam filter, even though both are based on an AI system. Indeed, this example alone illustrates the futility of lawmakers considering a general Artificial Intelligence Act, regulating the whole phenomenon top-town, administered by an Artificial Intelligence Agency.

Accordingly, there is no need for one all-encompassing definition for "algorithms" and "AI". Rather, it is more important to understand the different characteristics of various algorithms and AI applications and how they are used in practice.

The same applies for the term "robot", for which no universally valid definition has yet emerged.[21] Admittedly, at the international level some definitions can be found. For example, the International Standards Organization defines a robot as an "actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks".[22] This interpretation, however, is a functional rather than legal definition for the purpose of technical standards. Ultimately, all attempts at providing an encompassing definition are a fruitless exercise because of the extremely diverse nature of robots, ranging from driverless cars, prosthetic limbs, orthotic exoskeletons, and manufacturing (industrial) robots to care robots, surgical robots, lawn mowers, and vacuum cleaners. Instead of finding a common definition, greater insight can be gained from keeping all these robots separate, looking at the peculiarities and differences between them.

For our purposes, it is therefore sufficient to use a broad definition according to which a robot is a machine that has a physical presence, can be programmed, and has some level of autonomy depending inter alia on the AI algorithms used in such a system, or is, in short: "AI in action in the physical world".[23]

As there is no universally accepted characterization so far, this chapter uses the terms AI/algorithmic/self-learning/intelligent/smart/autonomous and/or robotic systems/machines interchangeably in order to refer to AI driven systems with a high degree of automation.

2. The Rise of Learning Algorithms

A particularly important subfield of AI is machine learning (ML). Instead of programming machines with specific instructions to accomplish particular tasks, ML algorithms enable computers to learn from "training data", and even improve themselves without being explicitly programmed. Although the idea

[19] *Scherer*, Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies, (2016) 29(2) Harvard Journal of Law and Technology 353, at pp. 359 et seq. Cf. also *Lea*, Why we need a legal definition of artificial intelligence, The Conversation, September 2, 2015, http://theconversation.com/why-we-need-a-legal-definition-of-artificial-intelligence-46796.

[20] *Jabłonowska/Kuziemski/Nowak/Micklitz/Pałka/Sartor*, Consumer law and artificial intelligence. Challenges to the EU consumer law and policy stemming from the business's use of artificial intelligence. Final report of the ARTSY project, European University Institute (EUI) Working Papers, LAW 2018, 11, p. 4.

[21] By contrast, the EU Parliament calls for a uniform, Union-wide definition of robots in its 2017 resolution; *European Parliament*, Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, P8_TA(2017)0051. Critical *Lohmann*, Ein europäisches Roboterrecht – überfällig oder überflüssig?, (2017) Zeitschrift für Rechtspolitik (ZRP) 168, at p. 169.

[22] ISO 8373, 2012, available at https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en. Additionally, ISO makes a distinction between industrial robots and service robots, as well as between personal service robots and service robots for personal use.

[23] Cf. *AI HLEG*, A Definition of AI (n. 17), p. 4.

of creating "learning machines" was already present in the early AI years,[24] only recent developments have brought algorithms to a new level, leading to an AI spring that outshines all the previous ones.

Over the years, ML developed into a number of different directions. By and large, they can be classified into three broad categories, depending on their learning pattern: supervised, unsupervised, and reinforcement learning.[25]

In a *supervised learning* setting, the algorithm uses a sample of labeled data to learn a general rule that maps inputs onto outputs.[26] For example, when the algorithm must learn how to recognize cats, the developer would give the system many examples of pictures of cats and the corresponding interpretation (that is, whether a cat is or is not in that picture). After the learning period, the system, through its ML algorithm, will then be able to generalize to know also how to interpret pictures of cats never seen before.

In an *unsupervised learning* setting, on the other hand, the algorithm attempts to identify hidden structures and patterns from unlabeled data.[27] This learning method is especially useful if data is rather unstructured. It can also be used in order to build better supervised learning algorithms, e.g. by combining the multitude of pixels of a picture into a small number of important recognizable features (such as the structures of eyes, nose, mouth), which in turn can then serve as an input for a supervised learning facial recognition algorithm.

Finally, in the *reinforcement learning* approach, the algorithm is not told how to "behave", but must learn in an (unknown but fixed) environment which actions yield the best (scalar) reward.[28] ML applications based on this approach are used especially in a dynamic environment, such as driving a vehicle or playing a game (as, for example, DeepMind's AlphaGo).

III. Overview

Autonomous systems, especially those based on machine learning, pose a number of legal and ethical problems (cf. C-G.). Before going into these questions in detail, it is worth taking a broader look at the general characteristics of algorithmic systems, which are ultimately responsible for the irritations and disruptive effects we are currently observing worldwide in all legal systems (B.).

[24] The idea of "learning machines" was raised as early as 1950 by *Turing*, Computing Machinery and Intelligence, Mind, (1950) LIX(236) A Quarterly Review of Psychology and Philosophy 433, at p. 456 (suggesting that machines could simulate the child-brain which is "subjected to an appropriate course of education"). Just a few years later, in 1952, *Samuel* would then go on to create the first computer learning program, a Checkers-playing program which improved itself through self-play; *Samuel*, Some Studies in Machine Learning Using the Game of Checkers, (1959) 3 IBM Journal of Research and Development 210.

[25] *Anitha/Krithka/Choudhry*, (2014) 3(12) International Journal of Advanced Research in Computer Engineering & Technology 4324, http://ijarcet.org/wp-content/uploads/IJARCET-VOL-3-ISSUE-12-4324-4331.pdf; *Buchanan/Miller*, Machine Learning for Policymakers. What It Is and Why It Matters, Harvard Kennedy School, Belfer Center for Science and International Affairs, Paper, June 2017; *Mohri/Rostamizadeh/Talwalkar*, Foundations of Machine Learning, 2012. Cf. also, in this book, *Haddadin/Knobbe*, Chapter 1.

[26] *Anitha/Krithka/Choudhry* (n. 25), 4325 et seq.

[27] *Anitha/Krithka/Choudhry* (n. 25), 4328 et seq.

[28] For a comprehensive introduction to reinforcement learning see *Sutton/Barto*, Reinforcement Learning – An Introduction, 2017.

B. The Problematic Characteristics of AI Systems from a Legal Perspective

I. Complexity and Connectivity

Some of these characteristics are already known in connection with other IT systems, especially the complexity and connectivity of computer systems: The increasing interconnectivity of computers leads to a multiplicity of actions and actors. This applies in particular to smart objects in the Internet of Things (IoT). The individual consumer who acquires a smart object is regularly confronted with a large number of potential contractual partners who owe various services (hardware, digital content, digital services, end user license agreements with third parties), all of which are required together for the IoT to function properly.[29] As a result, it is often no longer clear to the individual with whom she has concluded a contract. Moreover, there is a serious problem of proof: Although the purchaser cannot always ascertain why her product does not work (i.e. whether it is due to hardware or digital content), the burden of proof for the existence of a defect lies in principle with her, so that she is also burdened with the costs of determining its cause.

It can also be the case that the individual AI system works flawlessly on its own and does not exhibit any problematic behavior at all, but that a functional failure and/or damage occurs only through the interaction of different software agents. Some consider the so-called Flash Crash on May 6, 2010[30] to be just such an event: $1 trillion in market value vanished in less than an hour, and trading had to be suspended. When such an event occurs, assumptions are destroyed about the individuality of actors who are constitutive in the attribution of action and responsibility. Both the actor and the causal relationships are difficult, if not impossible, to identify.

In order to address these problems, various solutions have been proposed. For contractual claims it is discussed whether the doctrine of privity of contract must be overcome, for example by accepting linked contracts[31] or the concept of a contractual network.[32] For non-contractual claims, some scholars propose a pro-rata liability for all those involved in the network, requiring actors themselves to stand up for the unlawful behavior of the networked algorithms,[33] whereas others are in favor of attributing legal responsibility not to people, organizations, networks, software agents, algorithms, but rather to risk pools and the decisions themselves.[34]

---

[29] *Wendehorst*, Sale of Goods in the Digital Age – From Bipolar to Multi-party Relationships, in: UNIDROIT (ed.), Eppur si muove: The Age of Uniform Law. Essays in honour of Michael Joachim Bonell to celebrate his 70th birthday, Vol. 2, 2016, pp. 1873-1887.

[30] *Commodity Futures Trading Commission/Securities & Exchange Commission* (2010), Findings Regarding the Market Events of May 6, 2010, Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues, https://www.sec.gov/news/studies/2010/marketevents-report.pdf. See also *Kirilenko/Kyle/Samadi/Tuzun*, The Flash Crash: High-Frequency Trading in an Electronic Market, (2017) Journal of Finance, https://ssrn.com/abstract=1686004 or http://dx.doi.org/10.2139/ssrn.1686004.

[31] *Forgó*, in: Forgó/Zöchling-Jud, Das Vertragsrecht des ABGB auf dem Prüfstand: Überlegungen im digitalen Zeitalter, Gutachten Abteilung Zivilrecht, Verhandlungen des zwanzigsten österreichischen Juristentages Salzburg, 2018, pp. 276 et seq.

[32] Cf. *Cafaggi*, Contractual Networks and the Small Business Act: Towards European Principles?, EUI Working Paper Law no. 2008/15, http://cadmus.iue.it/dspace/bitstream/1814/8771/1/LAW_2008_15.pdf; *Idelberger*, Connected Contracts Reloaded – Smart Contracts As Contractual Networks, in: Grundmann (ed.), European Contract Law in the Digital Age, 2018, pp. 205 et seq.

[33] *Spiecker*, Zur Zukunft systemischer Digitalisierung – Erste Gedanken zur Haftungs- und Verantwortungszuschreibung bei informationstechnischen Systemen, (2016) Computer und Recht (CR) 698, at p. 703.

[34] *Teubner*, Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten, (2018) 218 Archiv für die civilistische Praxis (AcP) 155.

II. From Causation to Correlation

Another characteristic of AI systems in the context of Big Data analysis is a shift "from causation to correlation".[35] Most data mining techniques rely on inductive knowledge and correlations identified within a dataset. Instead of searching for causation between the relevant parameters, powerful algorithms are used to spot patterns and statistical correlations.[36]

Relying on correlations when statistical analysis indicates a significant relationship between factors provides clear benefits in terms of speed and costs.[37] However, it becomes problematic when correlation is increasingly seen as sufficient grounds for directing action without first establishing causality. Data analysis, actions, and far reaching decisions (e.g. scoring values or a medical diagnosis) relying on mere correlations in probability values might be severely flawed: First and foremost, relying on correlations without investigating causal effects bears the risk that correlations are "forced" on the data.[38] As *Marcus & Davis* explain, big data detecting correlations "never tells us which correlations are meaningful. A big data analysis might reveal, for instance, that from 2006 to 2011 the United States murder rate was well correlated with the market share of Internet Explorer: Both went down sharply. But, it's hard to imagine there is any causal relationship between the two."[39] Moreover, even if a strong statistical correlation is found, this only says something about a particular (sub)group of persons, but not about the individual belonging to that (sub)group. Finally, pure correlation statements do not allow individuals to engage in self-improvement. How, for example, should a policyholder behave if she is informed that she has come into a worse tariff not because her driving is risky, but because a Big Data analysis has shown that her Facebook "likes" indicate an increased accident risk? Thus, finding causation can be crucial in promoting the quality of the entire process and ensuring that in the end individuals are treated fairly.

III. Autonomy

Probably the biggest problem is the growing degree of autonomy of AI systems and smart robotics.[40] Self-learning systems are not explicitly programmed; instead, they are trained by thousands and

---

[35] *Mayer-Schönberger/Cukier*, Big Data: A Revolution that will Transform How We Live, Work and Think, 2013, p. 14, 15, 18, and p. 163: "Big Data does not tell us anything about causality".

[36] Some commentators believe that new data-mining techniques will free science of the constraints of theory, establishing a world in which the search for causation will no longer be paramount as correlation takes the center stage. Chris Anderson refers to this phenomenon as "the end of theory"; *Anderson*, The End of Theory, Wired, July 2008, at p. 108. Critically, *Skopek*, Big Data's Epistemology and Its Implications for Precision Medicine and Privacy, in: Cohen/Fernandez Lynch/Vayena/Gasser (eds.), Big Data, Health Law, and Bioethics, 2018, pp. 30 et seq.

[37] *Zarsky*, Correlation versus Causation in Health-Related Big Data Analysis. The Role of Reason and Regulation, in: Cohen/Fernandez Lynch/Vayena/Gasser (eds.), Big Data, Health Law, and Bioethics, 2018, p. 42, 50.

[38] *Silver*, The Signal and the Noise. Why So Many Predictions Fail – but Some Don't, 2012, p. 162.

[39] *Marcus/Davis*, Eight (No, Nine!) Problems With Big Data, N.Y. Times, April 6, 2014, http://www.nyti.ms/1kgErs2. Cf. also *Kosinski/Stillwell/Graepel*, Private traits and attributes are predictable from digital records of human behavior, (2013) Proceedings of the National Academy of Sciences of the United States of America (PNAS) 5802, http://www.pnas.org/content/110/15/5802.full, stating a correlation between high intelligence and Facebook likes of "thunderstorms", "The Colbert Report", and "curly fries", while users who liked the "Hello Kitty" brand tended to be higher in openness and lower in conscientiousness, agreeableness, and emotional stability.

[40] In the discussion, various criteria are offered as the starting point from which an AI system can be regarded as autonomous. What is clear, however, is that autonomy seems to be a gradual phenomenon. On the different concepts of autonomy cf. *Bertolini*, Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules, (2013) 5(2) Law, Innovation and Technology 214, at pp. 220 et seq.; *Floridi/Sanders*, On the

millions of examples, so that the system develops by learning from experience. The increasing use of ML systems poses great challenges for legal systems. With a certain degree of automation it seems impossible to ascertain with certainty whether the programmer, the producer, or the operator is responsible for actions caused by such systems. Specific problems arise in particular from the point of view of foreseeability and causation.

As for foreseeability, we have already seen numerous instances of AI making decisions that a person would not have made or would have made differently. A particularly fascinating example highlighted by *Mathew Scherer*[41] comes from C-Path, a machine learning program for the detection of cancer. Pathologists believed that the study of tumor cells is the best method for diagnosing cancer, whereas studying the supporting tissue (stroma) might only aid in cancer prognosis. But in a large study, C-Path found that the properties of stroma were actually a better prognostic indicator for breast cancer than the properties of the cancer cells themselves – a conclusion that contradicted both common sense and predominant medical thinking.[42] Another example concerns AlphaGo, a computer program developed by Google DeepMind that defeated Lee Sedol, the South Korean world champion Go player, in a five-game match in March 2016. As DeepMind noted on their blog, "during the games AlphaGo played a handful of highly inventive winning moves, one of which — move 37 in game two — was so surprising it overturned hundreds of years of received wisdom and has been intensively examined by players since. In the course of winning, AlphaGo somehow taught the world completely new knowledge about perhaps the most studied game in history."[43] Both examples show that AI systems may act in unforeseeable ways, as they come up with solutions that humans may not have considered, or that they considered and rejected in favor of more intuitively appealing options.

The experiences of a self-learning AI system can also be viewed, as *Scherer* correctly points out, as a superseding cause – that is, "an intervening force or act that is deemed sufficient to prevent liability for an actor whose tortious conduct was a factual cause of harm"[44] – of any harm that such systems cause. This is especially true when an AI system learns not only during the design phase, but also after it has already been launched on the market. In this case, even the most cautious designers, programmers, and manufacturers will not be able to control or predict what an AI system will experience in the environment.

For all these reasons, self-learning systems with a high degree of automation cause considerable irritations in legal systems.[45]

Morality of Artificial Agents, in: Anderson/Anderson (eds.), Machine Ethics, 2011, pp. 184 et seq., at p. 192; *Zech*, Zivilrechtliche Haftung für den Einsatz von Robotern: Zuweisung von Automatisierungs- und Autonomierisiken, in: Gless/Seelmann (eds.), Intelligente Agenten und das Recht, 2016, 163, at pp. 170 et seq., fn. 16. For the different levels of automation for self-driving cars, see the categories proposed by SAE International (Society of Automotive Engineers) and DOT (US Department of Transportation); *DOT*, Federal Automated Vehicles Policy, September 2016, p. 9, available at https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety.

[41] *Scherer*, (2016) 29(2) Harvard Journal of Law and Technology 353, at p. 363 und p. 364.

[42] *Beck* et al., Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival, (2011) 108(3) Sci. Translational Med. 1.

[43] *Hassabis*, The mind in the machine: Demis Hassabis on artificial intelligence, Financial Times, April 21, 2017, https://www.ft.com/content/048f418c-2487-11e7-a34a-538b4cb30025.

[44] Restatement (Third) of Torts: Phys. & Emot. Harm § 34 cmt. b (AM. LAW INST. 2010).

[45] Cf. *infra*, E.

IV. Algorithms as Black Boxes

A particular concern in relation to advanced ML techniques is the opacity of many algorithmic-decision-making (ADM) systems. The notion of black-box AI refers to such scenarios, where we can see only input data and output data for algorithm-based systems without understanding exactly what happens in between.[46]

Explainability is relevant for a number of reasons.[47] For a researcher or developer, it is crucial to understand how their system or model is working in order to debug or improve it. For those affected by an algorithmic decision, it is important to comprehend why the system arrived at this decision in order to understand the decision, develop trust in the technology, and – if the ADM process is illegal – initiate appropriate remedies against it. Last but not least, explainability enables experts (and regulators) to audit ADM and verify whether legal regulatory standards have been complied with.

According to *Gunning*[48] and *Waltl/Vogl*,[49] an ADM system has a high degree of explainability if the following questions can be answered:

- Why did that output happen?
- Why not some other output?
- For which cases does the machine produce a reliable output?
- Can you provide a confidence score for the machine's output?
- Under which circumstances, i.e. state and input, can the machine's output be trusted?
- Which parameters effect the output most (negatively and positively)?
- What can be done to correct an error?

In order to answer these questions, it is helpful to distinguish the following three dimensions that can be found in every ADM system: the process level, the model level, and the classification level:[50]

- The *process level* refers to the different steps an AI system has gone through in order to make an autonomous decision, usually beginning with the data acquisition phase; followed by data pre-processing; the selection of features; the training and application of the AI model; and the post-processing phase, in which steps are taken to improve and revise the output of the AI model. The exact knowledge of these steps is necessary to understand decisions. If, for example, a discriminatory decision is based on biased training data, precise knowledge of the data acquisition phase is required.
- The *model level*, on the other hand, refers to the different types of algorithms that are used for decision making, e.g. decision trees, Bayesian networks, support vector machines, k-nearest neighbors, or neural networks.
- This must be distinguished from the *classification level*, which provides information about which attributes (e.g. gender, age, salary) are used in the model and what weight is given to each attribute.

[46] Additionally, it might be that the inputs themselves are entirely unknown or known only partially.

[47] *Anand et al*., Effects of Algorithmic Decision-Making and Interpretability on Human Behavior: Experiments using Crowdsourcing, 2018, www.l3s.de/~gadiraju/publications/HCOMP18.pdf.

[48] *Gunning*, Explainable Artificial Intelligence (XAI), 2017, https://www.darpa.mil/attachments/XAIProgramUpdate.pdf.

[49] *Waltl/Vogl*, Explainable Artificial Intelligence – the New Frontier in Legal Informatics, Jusletter IT, February 22, 2018.

[50] *Waltl/Vogl*, Increasing Transparency in Algorithmic Decision-Making with Explainable AI, (2018) Datenschutz und Datensicherheit (DuD) 613.

Opacity in ML algorithms can have quite different causes.[51] First, it might be that algorithms are kept secret intentionally for the sake of competitive advantage,[52] national security,[53] or privacy.[54] Keeping an AI system opaque can also be important to ensure its effectiveness, as in preventing spambots from using the disclosed algorithm to attack the system.[55] Moreover, it might be that corporations protect their ADM system to avoid or confound regulation, and/or to conceal manipulation or discrimination of consumers.[56] Secondly, opacity can be an expression of technical illiteracy. Writing and reading code as well as designing algorithms requires expertise that the majority of the population does not have. Thirdly, it may be that opacity arises due to an unavoidable complexity of ML models. As *Burrell* notes, in the era of Big Data, "Billions or trillions of data examples or tens of thousands of properties of the data (termed "features" in machine learning) may be analyzed. (…) While datasets may be extremely large but possible to comprehend, and code may be written in clarity, the interplay between the two in the mechanism of the algorithm is what yields the complexity (and thus opacity)."[57]
Apart from that, it is important to understand that different classes of ML algorithms have different degrees of transparency as well as performance.[58] Thus, for example, deductive and rule-based systems (such as decision trees) have a high degree of transparency: since each node represents a decision, the way to the respective leaf can be understood as an explanation for a concrete decision. By comparison, artificial neural networks (ANN), especially deep learning systems, show a very high degree of opacity. In such a network, all learned information is not stored at a single point but is distributed all over the neural net by modifying the architecture of the network and the strength of individual connections between neurons (represented as input "weights" in artificial networks). Therefore, ANN systems possess a high degree of unavoidable complexity and opacity. On the other hand, when it comes to performance, it is precisely ANNs that show a much higher degree of accuracy and effectiveness than decision trees.[59]
We are therefore faced with a dilemma: How can human-interpretable systems be designed without sacrificing performance?

C. Fundamental Questions

The use of AI systems and smart robots – in addition to the problems discussed above – raises a number of fundamental questions.

I. Replacement of Humans By Machines: To What Extent?

Arguably the first and most fundamental question is to what extent we, as a society, are willing to replace humans with machines. This question arises in many areas, but above all when decisions are no longer made by people: When should a human decision be replaced with an algorithm? Which decisions should in any case be made by a human being? Are there certain decisions that must always

[51] *Burrell*, How the machine "thinks": Understanding opacity in machine learning algorithms, (2016) January-June Big Data & Society, 1.
[52] *Kitchin*, (2017) 20(1) Information, Communication & Society 14.
[53] *Leese*, The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union, (2014) 45(5) Security Dialogue 494.
[54] *Mittelstadt/Allo/Taddeo/Wachter/Floridi*, The ethics of algorithms: Mapping the debate, (2016) July-September Big Data & Society 1, at p. 6.
[55] *Sandvig/Hamilton/Karahalios/Langbort*, Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms, in: Annual Meeting of the International Communication Association, 2014, http://social.cs.uiuc.edu/papers/pdfs/ICA2014-Sandvig.pdf, pp. 1 et seq., at p. 9.
[56] *Pasquale*, The Black Box Society: The Secret Algorithms that Control Money and Information, 2015, p. 2.
[57] *Burrell* (n. 51) at p. 5.
[58] *Waltl/Vogl*, Explainable Artificial Intelligence – The New Frontier in Legal Informatics, Jusletter IT, February 22, 2018.
[59] *Waltl/Vogl* (n. 58).

be made by humans for deontological or other (ethical/legal) reasons? To what extent should an algorithm be able to influence a human decision?

Such questions are currently being discussed, particularly with regard to the use of lethal autonomous weapon systems (LAWS): Is it right for machines to have the power of life and death over humans or the ability to inflict serious injury? Are LAWS both inherently unethical and unlawful under current international humanitarian law? Do we need a new international agreement?[60] – The consensus seems to be that the decision to kill a human person in a concrete combat situation cannot be delegated to a machine.[61]

The question as to whether decisions should be delegated to machines also arises in many other cases, especially when state decisions are involved:

- To what extent can administrative decisions be automated?[62] Is the idea of algorithmic regulation in line with the nondelegation doctrine, the principles of procedural due process, equal protection, and/or the principles of reason-giving and transparency?
- How far should the judiciary go in using AI systems to resolve a dispute or as a tool to assist in judicial decision-making?[63] How can we ensure that the design and implementation of AI tools and services in the judicial system are compatible with fundamental rights, especially the guarantees of the right of access to the judge, the right to a fair trial (equality of arms and respect for the adversarial process), and the rule of law?
- What are the advantages and drawbacks of legal automation?[64] How can the law govern human behavior through codes, IT architectures, and design? Should legislators be allowed to adopt "personalized laws" by tailoring laws/legal provisions to the individual needs and characteristics of addressees?[65]

How about the private sector? To what extent may private companies delegate decisions to an algorithm and which decisions should be reserved for humans alone?[66] How do ADM procedures

[60] *Melzer*, Targeted killing in international law, 2008; *Wagner*, The dehumanization of international humanitarian law: legal, ethical, and political implications of autonomous weapons systems, (2014) 47 Vanderbilt J Transnatl Law 1371; *Crawford*, The principle of distinction and remote warfare, (2016) Sydney Law School Research Paper No. 16/43; *Ohlin*, Remoteness and reciprocal risk, (2016) Cornell Legal Studies Research Paper No. 16-24.

[61] *European Parliament*, Resolution of 12 September 2018 on autonomous weapon systems, P8_TA-PROV(2018)0341; *Scharre*, The trouble with trying to ban "killer robots", World Economic Forum, September 4, 2017, https://www.weforum.org/agenda/2017/09/should-machines-not-humans-make-life-and-death-decisions-in-war/.

[62] Cf. *Coglianese/Lehr*, Regulating by Robot: Administrative Decision Making in the Machine-Learning Era, (2017) 105 Georgetown Law Journal 1147; https://ssrn.com/abstract=2928293.

[63] Cf. *Council of Europe*, European Commission for the Efficiency of Justice (CEPEJ), European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, adopted by the CEPEJ during its 31st Plenary meeting (Strasbourg, 3-4 December 2018), CEPJ(2018)14 (*Council of Europe*, Ethical Charter).

[64] *Pagallo/Durante*, The Pros and Cons of Legal Automation and its Governance, (2016) 7 European Journal of Risk Regulation 323.

[65] *Porat/Strahilevitz*, Personalizing Default Rules and Disclosure with Big Data, (2014) 112 Michigan Law Review 1417; *Ben-Shahar/Porat*, Personalizing Negligence Law, (2016) 91 NYU Law Review 627; *Hacker*, Personalizing EU Private Law. From Disclosures to Nudges and Mandates, (2017) 25 European Review of Private Law (ERPL) 651. Moreover, see the special issue of the University of Chicago Law Review Vol. 86, No. 2, March 2019, on "Personalized Law".

[66] *Möslein*, Robots in the Boardroom: Artificial Intelligence and Corporate Law, in: Barfield/Pagallo (eds.), Research Handbook on the Law of Artificial Intelligence, 2018, https://ssrn.com/abstract=3037403.

impact consumers' autonomy and freedom to make decisions, as well as how they access products and services?[67]

At present, there is no legal system in the world that provides satisfactory answers to these questions. In the European Union, Art. 22 GDPR[68] prohibits fully automated decisions. However, this provision has a rather limited scope of application. First, it establishes numerous exceptions in Art. 22(2) GDPR. And secondly, it only covers decisions "based *solely* on automated processing" of data (Art. 22(1) GDPR). Since most algorithmically prepared decisions still involve a human being, the majority of ADM procedures is not covered by the prohibition of Art. 22 GDPR.[69]
The policy decision as to which decisions must be reserved for humans is by no means an easy one,[70] as the transfer of decision-making power to machines brings great advantages, especially in terms of efficiency and costs. The political decision *not* to transfer certain tasks to machines can thus lead to economic loss. Moreover, in most cases it is impossible to make a clear distinction between purely machine and purely human decisions. Rather, many decisions are made in a more or less symbiotic relationship between humans and machines. For this reason, it is very difficult to determine at what point in this continuum the "essence of humanity" is compromised.

II. Brain-Computer Interfaces (BCI) and Human Enhancement

An equally fundamental question is to what extent the use of BCI should be permitted. This problem arises in particular when a healthy person connects his body with a BCI in order to be more efficient (BCI enhancement). The blurring of the distinction between man and machine makes it more difficult to assess the limits of the human body and raises questions concerning free will and moral responsibility.[71]

Should everyone be free to expand and influence their cognitive, mental, and physical abilities beyond the boundaries of the natural? Is such a fusion socially desirable and ethically acceptable? If we restrict

---

[67] Cf. *infra*, G.I.4.

[68] GDPR Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) O.J. 2016 L 119/1.

[69] It is still unclear which type of human participation deprives a decision of its automated nature. Art. 29 Working Party (WP) argues that a decision cannot be regarded as wholly automated if an automated profile is accompanied by an "additional meaningful intervention carried out by humans before any decision is applied to an individual"; Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, adopted on 3 October 2017, as last Revised and Adopted on 6 February 2018, WP251rev.01, p. 8. *Bygrave* argues that decisions formally attributed to humans but originating "from an automated data-processing operation the result of which is not actively assessed by either that person or other persons before being formalised as a decision," would fall under the scope of 'automated decision-making', *Bygrave*, Automated Profiling: Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling, (2001) 17 Computer Law & Security Review 17. However, as *Wachter/Mittelstadt/Floridi* correctly point out, whereas the EP's proposed amendments suggested the words "*based solely or predominantly on automated processing*", the final text did not adopt the word "predominantly", suggesting that a strict reading of "solely" was intended; *Wachter/Mittelstadt/Floridi*, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, (2017) 7(2) International Data Privacy Law 76, at p, 92. The EP Amendments are available at: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-%2F%2FEP%2F%2FTEXT%2BREPORT%2BA7-2013-0402%2B0%2BDOC%2BXML%2BV0%2F%2FEN&language=EN.

[70] *Burri*, Künstliche Intelligenz und internationales Recht, (2018) Datenschutz und Datensicherheit (DuD) 603, at pp. 606 et seq.

[71] *Schermer*, The mind and the machine. On the conceptual and moral implications of brain-machine interaction, (2009) 3(3) Nanoethics 217, http://dx.doi.org/10.1007/s11569-009-0076-9.

individual enhancement, should those limits include only biological considerations (in order to restore the body to a "normal" state) or psychological ones as well? Does our existing liability framework provide appropriate remedies for those who suffer injuries caused by BCI systems, especially in cases where users may be able to send thoughts or commands to other people including unintended commands? Is the existing data protection law sufficient or do we need to protect highly sensitive personal BCI data emanating from the human mind in a particular way? What precautions must be taken against brain spyware?

Leading international neuroscientists see us facing such questions. They demand ethical and legal guidelines for the use of BCI.[72]

D. Safety and Security Issues

The use of AI and smart robotics also raises a number of safety and security issues.

I. Superintelligence as a Safety Risk?

The AI safety problem is often associated with the concern that a "superintelligence" – or Artificial General Intelligence (AGI) – will inevitably turn against humanity and trigger a "post-human" future.[73] In order to address this concern, various (global and local) solutions have been proposed,[74] especially (i) "No AI" solutions consisting of an international ban on AI, legal/technical relinquishment, destruction of the capability to produce AI, and a slowdown of AI creation; (ii) the "One AI" solution in which the first AI will become dominant and prevent the development of other AIs; (iii) "Many AI" solutions in which a network of AIs may provide global safety; and (iv) solutions in which "humans are incorporated inside AI".

Such discussions, however, ultimately lead in the wrong direction. Not only is it controversial among experts whether superintelligence will ever happen[75] and whether – once created – it might do something dangerous.[76] What's more, the ongoing discussion about a rising superintelligence obscures our view of the actual safety and security problems we are facing today.

---

[72] *Clausen* et al., Help, hope, and hype: Ethical dimensions of neuroprosthetics, (2017) Science, Vol. 356, Issue 6345, pp. 1338 et seq., http://science.sciencemag.org/content/356/6345/1338. *Bostrom/Sandberg*, Cognitive Enhancement: Methods, Ethics, Regulatory Challenges, (2009) 15 Sci Eng Ethics 311; *Holder* et al., Robotics and law: Key legal and regulatory implications of the robotics age (part II of II), (2016) 32 Computer Law & Security Review 557, at pp. 570 et seq.

[73] *Bostrom*, Superintelligence, 2014; *Russell*, 3 principles for creating safer AI, 2017, retrieved from https://www.youtube.com/watch?v=EBK-a94IFHY; *Yudkowsky*, Artificial Intelligence as a Positive and Negative Factor in Global Risk, in: Cirkovic/Bostrom (eds.), Global Catastrophic Risks, 2008.

[74] *Turchin/Denkenberger*, Classification of the Global Solutions of the AI Safety Problem. PhilArchive copy v1: https://philarchive.org/archive/TURCOT-6v1; *Sotala*/*Yampolskiy*, Responses to catastrophic AGI risk: A survey, last modified September 13, 2013, intelligence.org/files/ResponsesAGIRisk.pdf.

[75] According to a survey by *Müller/Bostrom*, which gathered opinions from the world's top 100 most cited AI researchers, the median estimate for the time of emergence of what might be labelled human-level AI is 2050, with experts forecasting the emergence of superintelligence by the turn of the century, *Müller/Bostrom*, Future progress in artificial intelligence: A survey of expert opinion, in: Müller (ed.), Fundamental Issues of Artificial Intelligence, 2016, pp. 553 et seq. According to the survey by Grace et al., there is a "50% chance AI will outperform humans in all tasks in 45 years"; *Grace/Salvatier/Dafoe/Zhang/Evans*, When Will AI Exceed Human Performance? Evidence from AI Experts, last revised May 3, 2018, arXiv:1705.08807.

[76] Cf. *Häggström*, Remarks on Artificial Intelligence and Rational Optimism, in: European Parliament (ed.), Should we fear artificial intelligence?, March 2018, PE 614.547, pp. 19 et seq., at p. 21.

II. Current Safety Risks

First of all, one might wonder whether existing (product) safety rules are sufficient to ensure an adequate level of safety. Special safety requirements exist above all in the field of robotics. The ISO and IEC standards governing robot safety include:

- Industrial robots, ISO 10218-1 and ISO 10218-2:2011
- Personal care robots, ISO 13482:2014
- Collaborative robots, ISO/TS 15066:2016
- Robotic lawn movers, IEC 60335-2-107
- Surgical robots, IEC 80601-2-78
- Rehabilitation robots, IEC 80601-2-77

In Europe, these safety requirements are transferred into national law by the EU Machinery Directive 2006/42. Whether the international standards are fit to deal with innovative robots with machine intelligence is highly controversial. The International Federation of Robotics believes that existing safety standards are sufficient to cover current developments in the use of AI in robots in commercial applications, and that no additional regulation is required.[77] By contrast, the European Commission's evaluation report of the Machinery Directive is more cautious, highlighting that the suitability of the Directive may be tested when it comes to AI-powered advanced robots and autonomous self-learning systems.[78] In the same vein, the UK Science and Technology Committee maintains that so far, according to experts, "no clear paths exist for the verification and validation of autonomous systems whose behavior changes with time".[79] Another report notes that regulation lags behind and is not yet consolidated, resulting in gaps and overlaps between standards.[80]

International standard setting organizations also see a need for action. The work in this domain has already started in the Joint Technical Committee 1 between ISO and IEC (JTC 1) and its subcommittee (SC) 42 (JTC 1/SC 42)[81] under the lead of the American National Standards Institute (ANSI)[82] and US secretariat. Similar initiatives have also been taken by the European standardization organizations CEN and CENELEC since 2018.[83]

III. Security Risks Due to Malicious Use of AI

Security issues also play a crucial rule. AI is a dual-use technology that can be used both for beneficial and harmful ends, bringing enormous security risks to not only individuals, governments, industries, and organizations but also to the future of humanity. Malicious use of AI could, as a recent report suggests,[84] threaten physical security (e.g. non-state actors weaponizing consumer drones), digital

[77] *International Federation of Robotics*, Artificial Intelligence in Robotics, May 2018, https://ifr.org/downloads/papers/Media_Backgrounder_on_Artificial_Intelligence_in_Robotics_May_2018.pdf.
[78] Commission Staff Working Document, Evaluation of the Machinery Directive, SWD(2018) 161 final, p. 38.
[79] *UK Science and Technology Committee*, Robotics and artificial intelligence, Fifth Report, Session 2016-17, HC 145, https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf.
[80] *Jacobs*, Report on regulatory barriers, Robotics Coordination Action for Europe Two, Grant Agreement Number: 688441, March 3, 2017.
[81] https://www.iso.org/committee/6794475.html.
[82] https://www.ansi.org/.
[83] *Schettini Gherardini*, Is European standardization ready to tackle Artificial Intelligence?, September 19, 2018, https://www.linkedin.com/pulse/european-standardization-ready-tackle-artificial-bardo/.
[84] *Brundage* et al., The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, arXiv preprint arXiv:1802.07228, 2018. Cf. also *King* et al., Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions, (2019) Science and Engineering Ethics, https://doi.org/10.1007/s11948-018-00081-0.

security (e.g. through criminals training machines to hack), and political security (e.g. through privacy-eliminating surveillance, profiling, and repression, or through automated and targeted disinformation campaigns). As AI capabilities become more powerful and widespread, the authors of the report expect (i) an expansion of existing threats (because the costs of attacks may be lowered and AI might enable larger-scale and more numerous attacks), (ii) an introduction of new threats (by completing tasks that would be otherwise impractical for humans), and (iii) a change to the typical character of threats (because AI enables more effective, finely targeted, difficult-to-attribute attacks).

In light of these considerations, one key question for future regulation is: What safeguards should be put in place to prevent the malicious use of AI systems and smart robots? Are the existing security regulations sufficient or do we need new rules specifically tailored to the risks posed by AI?

E. Accountability, Liability and Insurance for Autonomous Systems

Closely related to these questions is the issue of accountability and liability for autonomous systems.

I. Emerging Questions

The use of semi-autonomous and autonomous systems leads to a loss of human control over the system and its "actions". With the increasing independence of technical systems, people's ability to influence technology is diminishing. The more complex the tasks assigned to machines, the greater the probability that the result will not correspond to the user's, the systems owner's/keeper's and/or the manufacturer's ideas and wishes.

This growing degree of autonomy inevitably raises the question of who is responsible if the autonomous AI system "makes" a declaration of intent to conclude a contract, "violates" a contractual obligation, or "commits" a wrong or even a crime. All major legal systems around the world are based on the premise that only natural and legal persons have legal capacity and are thus actors. From this anthropocentric perspective, technical artifacts are seen only as tools used by humans. It is precisely this perspective, however, that turns out to be problematic as the degree of autonomy of machines increases. With increasing automation, it becomes more and more difficult to determine a responsible person who can be identified as the author of declarations of intent, and to whom it is possible to assign responsibility in order to establish liability:

- Is it even possible to attribute a computer-generated declaration to a human if the person in question has no concrete idea what exactly the system will do?
- What happens if the software agent, like a "*falsus procurator*", misrepresents a third party as the principal?
- Who is liable to pay damages if a largely autonomous machine causes damage? The manufacturer of the machine who has originally developed the autonomous system? The operator who is actually running the system by providing the required data, overseeing possible machine learning processes and pushing necessary updates? The systems owner/keeper or the user of the autonomous system? Or does the injured person, in the end, have to bear the damage himself, since no responsible person can be found?
- Do we need special rules in contract and tort law in order to tackle the allocation problems caused by the use of autonomous systems?

II. Overview of Opinions

All these questions have triggered a lively debate in literature both in the US and in Europe.[85] The solutions proposed to overcome these difficulties vary widely.

For contract law, for example, consideration is being given to (i) modifying contract doctrine by relaxing the requirement of intentionality in contract-making, (ii) understanding computers as a mere tool or legal agents, (iii) denying validity to transactions generated by autonomous systems, and (iv) granting legal personhood to software agents.

A similarly broad spectrum of opinions exists in tort law. Here, the suggestions range from (i) applying or expanding existing doctrines, for example by treating AI systems as we would employees or other assistants, minors, or animals – or by drawing on the existing liability measures such as the guardian liability in France; (ii) revising product liability law; (iii) introducing new strict liability regimes; to, once again, (iv) granting legal personhood to software agents.

III. Revising (Product) Liability Law in the European Union

1. Product Liability Law

In the European Union, product liability has been fully harmonized in all Member States through the Product Liability Directive 85/374/EEC, which establishes a system of strict liability, i.e. liability without fault, for producers when a defective product causes physical or material damage to the injured person. Whether this Directive is sufficient to take into account the special features of AI systems and robots is controversial.

First of all, it is not clear whether the Directive, with its definition of "product",[86] also covers non-tangible AI software and especially cloud technologies. Secondly, there are problems with regard to the fact that the Directive only applies to products and not to services.[87] Companies providing services such as (real-time) data services, data access, data analytics tools, and machine learning libraries are therefore not liable under the Product Liability Directive[88] so that national (non-harmonized) law

---

[85] Cf. the extensive references in note 94. For the US-American discussion cf. moreover *Geistfeld*, A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation, (2017) 105(6) California Law Review 1611; *Hubbard*, "Sophisticated Robots": Balancing Liability, Regulation, and Innovation, (2014) 66(5) Florida Law Review 1803; *Karnow*, The Application of Traditional Tort Theory to Embodied Machine Intelligence, in: Calo/Froomkin/Kerr (eds.), Robot Law, 2016, pp. 51 et seq.; *Selbst*, Negligence and AI's Human Users, Boston University Law Review, forthcoming, https://www.ssrn.com/abstract=3350508. For the European discussion cf. https://www.ssrn.com/abstract=3350508; *Pagallo*, The laws of robots: crimes, contracts, and torts, 2013; *Ebers*, La utilización de agentes electrónicos inteligentes en el tráfico jurídico: ¿Necesitamos reglas especiales en el Derecho de la responsabilidad civil?, InDret 3/2016 http://www.indret.com/pdf/1245.pdf; *Ebers*, Autonomes Fahren: Produkt- und Produzentenhaftung, in: Oppermann/Stender-Vorwachs (eds.), Autonomes Fahren, 2017, pp. 93 et seq., https://ssrn.com/abstract=3192911; *Wagner*, Produkthaftung für autonome Systeme, (2017) 217 Archiv für die civilistische Praxis (AcP) 707. Cf. also, in this book, *Navas Navarro*, Chapter 5, and *Janal*, Chapter 6.

[86] According to Art. 2(1) Product Liability Dir., "product" means all movables even if incorporated into another movable or into an immovable. The Directive, however, is silent on whether movables need to be tangible. Given that Art. 2(2) explicitly includes an intangible item like electricity, this could mean that tangibility is not a relevant criterion in terms of the Directive. On the other hand, it could be argued that electricity is an exception which cannot be generalized.

[87] Cf. ECJ, 21.12.2011, case C-495/10 (*Dutrueux*), ECLI:EU:C:2011:869; Commission Staff Working Document, Evaluation of Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations, and administrative provisions of the Member States concerning liability for defective products, SWD(2018) 157 final, p. 7. Cf. also the failed proposal for a Council Directive on the Liability of Suppliers of Services, COM(90) 482 final, O.J. 1990 C 12/8. The new Digital Content Directive (DCD) does not change this either, as damages are left to national law; cf. Art. 3(10) DCD.

[88] Service providers could only be liable if they manufacture the *product* as part of their service; if they put their name, trade mark, or other distinguishing feature on the product; or if the they import the product into the EU. However, they do not incur any product liability for the *service* rendered by them.

decides whether the (strict) liability rules developed for product liability can be applied accordingly to services.
Thirdly, there is the problem that, under Art. 4 Product Liability Directive, the injured party must prove that the product was defective when it was put into circulation. This is precisely what is difficult with learning AI systems. Is an unintended autonomous behavior of an AI system or an advanced robot a defect? Can the producer invoke the so-called "development risks defence" admitted by Art. 7(e) of the Directive and claim an exemption from liability on the basis of the argument that he could not have foreseen that the product would not provide the safety a person could expect? How can a defect be proven at all,[89] if the product's behavior is changing over its lifetime through learning experiences, on which the manufacturer has no more influence after putting into circulation? And how about cyber security? Could software vulnerability (for instance, a cyber-attack, a failure to update security software, or a misuse of information) be considered a defect?
Finally, the question arises whether the definition of damages is adequately laid out in the Directive, since it does not cover all types of possible damages, especially with regard to the damages which can be caused by new technological developments, such as economic losses, privacy infringements, or environmental damages.
With these factors in mind, the European Commission is currently in the process of assessing whether the national and EU safety and liability frameworks are fit for purpose considering these new challenges, or whether any gaps should be addressed. By mid-2019, a report is to be drawn up on this subject, supplemented by a guidance on the interpretation of the Product Liability Directive in the light of technological developments, to ensure legal clarity for consumers and producers in the event of defective products.[90]

2. Beyond Product Liability Law

Beyond product liability law, the issue remains as to when other persons are liable, in particular the operator, the owner/keeper, or the user. As these persons do not usually act negligently due to the high degree of autonomy of the AI system,[91] they can only be held accountable if there is strict liability. However, such a liability regime is usually lacking. Many legal orders are based on the principle of fault liability and only have specific rules of strict liability which are not open to analogy.
The European Parliament suggested in its resolution of 16 February 2017 on "Civil Law Rules on Robotics" to introduce a system of registration for specific categories of advanced robots and to adopt a future legislative instrument that should be based either on strict liability or a risk management approach, in each case supplemented by an obligatory insurance scheme backed up by a fund in order to ensure that reparation can be made for damages in cases where no insurance cover exists.[92]
Which persons should be liable is left open by the European Parliament.[93] Instead, the resolution only emphasizes in general terms that, according to the risk management approach, the person liable should be the one who is able "to minimize risks and deal with negative impacts". Once the parties bearing the ultimate responsibility have been identified, "their liability should be proportional to the actual level of instructions given to the robot and of its degree of autonomy". According to the European Parliament, the greater a robot's learning capability or autonomy, and the longer a robot's training, the greater the responsibility of its trainer should be.

[89] According to *Borghetti*, How can Artificial Intelligence be Defective?, in: Lohsse/Schulze/Staudenmayer (eds.), Liability for Artificial Intelligence and the Internet of Things, 2019, pp. 63 et seq., at p. 71, "defectiveness is not an adequate basis for liability", because in most circumstances, "it will be too difficult or expensive to prove the algorithm's defect."
[90] *European Commission*, Communication "Coordinated Plan on Artificial Intelligence", COM(2018) 795 final, p. 8.
[91] *Selbst* (n. 85).
[92] *European Parliament*, Resolution (n. 21), Nos. 2, 53, 57, 58.
[93] Critical *Lohmann* (n. 21), at p. 170.

Overall, the European Parliament's proposals remain very vague. There is no detailed discussion of who should be liable under what conditions, nor does it take into account the numerous proposals discussed in scientific literature.

IV. A Specific Legal Status for AI and Robots?

Another option that has been discussed for some time to overcome the autonomy problem is the conferral of (limited) legal personhood for robots and AI systems.[94] This idea has recently been taken up by the European Parliament in its resolution of 16 February 2017, suggesting that the legislature should consider:

> ''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.''[95]

This proposal has been sharply criticized, including in an open letter from several "Artificial Intelligence and Robotics Experts" in April 2018[96] claiming that the creation of a legal status of an "electronic person" should be discarded from both a technical perspective and a normative, i.e. legal and ethical, viewpoint.

Indeed, the introduction of a legal personhood for AI systems and/or robots is problematic for several reasons. First, it is questionable how AI systems and/or robots can be identified at all. Should personhood be conferred to the hardware, the software, or some combination of the two? To make things worse, the hardware and software may be dispersed over several sites and maintained by different individuals. They might be copied, deleted, or merged with other systems at very low costs. Even if software agents and/or robots had to be registered in the future, there would be a number of cases in which the "acting" machine could not be identified as a person at all. The introduction of a specific legal status for machines would therefore by no means solve all liability problems.

The second problem is that the electronic agent would have to be equipped with its own assets in order to compensate victims. Such a solution raises, first of all, the question of who should make the assets available: The manufacturer? The operator? The keeper/owner or the user? All of them? Or the robot itself depending on the profit it makes? Additionally, it remains unclear how the relevant funds should be paid out in the event of damages. If strict liability were applied here, it is not clear what advantages the introduction of a legal personhood would bring over introducing a stricter tort law. All these considerations show that creating a legal personhood for machines does not seem economically very efficient, as the same purpose can be more easily achieved simply by introducing strict lability and/or requiring insurance.[97]

---

[94] *Solum*, Legal personhood for artificial intelligence, (1992) 70 North Carolina Law Rev. 1231; *Karnow*, Liability for distributed artificial intelligence, (1996) 11 Berkeley Technol. Law J. 147; *Allen/Widdison*, Can Computers Make Contracts?, (1996) 9 Harv. J. L. & Tech. 26; *Sartor*, Agents in Cyber law, in: Proceedings of the Workshop on the Law of Electronic Agents, CIRSFID (LEA02) Gevenini, 2002, p. 7; *Teubner*, Rights of Non-Humans? Electronic Agents and Animals as New Actors in Politics and Law, (2006) 33 J.L. & SOCY 497, 502; *Matthias*, Automaten als Träger von Rechten. Plädoyer für eine Gesetzesänderung, PhD thesis, Berlin 2007; *Chopra/White*, A Legal Theory For Autonomous Artificial Agents, 2011. For an overview of the different concepts cf. *Koops/Hildebrandt/Jaquet-Chiffelle*, Bridging the Accountability Gap: Rights for New Entities in the Information Society? (2010) 11(2) Minnesota Journal of Law, Science & Technology 497; *Pagallo*, Apples, oranges, robots: four misunderstandings in today's debate on the legal status of AI systems, (2018) Phil Trans. R. Soc. A376.

[95] *European Parliament*, Resolution (n. 21), No. 59.

[96] http://www.robotics-openletter.eu/.

[97] *Nevejans,* Citizens' Rights and Constitutional Affairs – Legal Affairs, European Civil Law Rules in Robotics. Study, European Union 2016 http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf, p. 15; *Keßler*, Intelligente Roboter – neue Technologien im Einsatz, (2017) MultiMedia und Recht (MMR) 593.

Last but not least, many fear that the agenthood of artificial agents could be a means of shielding humans from the consequences of their conduct.[98] Damages provoked by the behavior and decisions of AI systems would not be upon the manufacturers, keepers, etc. Instead, only AI systems would be liable. Moreover, there is the danger of machine insolvency: "Money can flow out of accounts just as easily as it can flow in; once the account is depleted, the robot would effectively be unanswerable for violating human legal rights."[99]
All in all, the decision to confer a legal personality on an autonomous system would most likely lead to more questions and problems than solutions.

F. Privacy, Data Protection, Data Ownership and Access to Data

I. The Interplay Between Data and Algorithms

The current success of AI systems is based not only on the accessibility of cheap, robust computational power and ever more sophisticated algorithms, but also – and above all – on the availability of large amounts of data.

The more data is available to a learning algorithm, the more it can learn. In a groundbreaking paper, *Banko/Brill* showed in 2001 that the amount of data used to train ML algorithms has a greater effect on prediction accuracy than the type of ML method used.[100] Or, as *Peter Norvig*, chief scientist at Google, puts it: "We don't have better algorithms than anyone else. We just have more data."[101] This is precisely one of the reasons why some of the most successful companies today are the ones that have the most data on which to train their algorithms.

The race for AI is particularly influenced by the network effects that are already known from the platform economy: The more users a company has, the more personal data can be collected and processed to train the algorithms. This in turn leads to better products and services, which results in more customers and more data. In view of these network effects, some fear that the market for AI systems will become oligopolistic with high barriers to entry.[102] According to *Pedro Domingos*: "Control of data and ownership of the models learned from it is what many of the twenty-first century's battles will be about – between governments, corporations, unions, and individuals."[103]

Considering these points, a number of very different questions arise: When should companies and governments be allowed to process personal data using Big Data Analysis? Is (European) data protection law compatible with Big Data and AI systems? Who "owns" personal and non-personal data? How can companies protect investments that flow into Big Data analysis? Should we recognize

---

[98] *Bryson/Diamantis/Grant*, Of, for, and by the people: the legal lacuna of synthetic persons, (2017) 23 Artif. Intell. Law 273.
[99] *Bryson/Diamantis/Grant*, (n. 98), at p. 288.
[100] *Banko/Brill*, Scaling to Very Very Large Corpora for Natural Language Disambiguation, paper presented at Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, 2001.
[101] *Norvig*, quoted by Scott Cleland, Google's "Infringenovation" Secrets, Forbes, October 3, 2011, https://www.forbes.com/sites/scottcleland/2011/10/03/googles-infringenovation-secrets/#78a3795430a6.
[102] *Mayer-Schönberger/Ramge*, Reinventing Capitalism in the Age of Big Data, 2018. Some critics point out that as few as seven for-profit institutions – Google, Facebook, IBM, Amazon, Microsoft, Apple, and Baidu in China – hold AI capabilities that vastly outstrip all other institutions; *Iyengar*, Why AI consolidation will create the worst monopoly in U.S. history, TechCrunch, August 24, 2016, https://techcrunch.com/2016/08/24/why-ai-consolidation-will-create-the-worstmonopoly-in-us-history/; *Quora*, What Companies Are Winning the Race for Artificial Intelligence?, Forbes, February 24, 2017, https://www.forbes.com/sites/quora/2017/02/24/what-companies-are-winning-the-race-for-artificial-intelligence/#7a5025eaf5cd.
[103] *Domingos*, The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World, 2015, p. 45.

a "data ownership" or "data producer's rights"? To what extent must competitors be given the opportunity to gain access to data from other companies?

II. Privacy, Data Protection and AI Systems

1. How AI Systems and Robots Threaten Privacy

AI systems challenge current understandings of privacy. Most AI technologies have a deleterious impact on the right to privacy. On the one hand, AI systems based on ML cannot work without data. On the other hand, without AI systems it would not be possible to "understand" many of the unstructured masses of data. In a nutshell: Personal data is increasingly both the source and the target of AI applications. Accordingly, AI technologies create strong incentives to collect and store as much additional data as possible in order to gain meaningful new insights. This trend is further reinforced by the shift to ubiquitous tracking and surveillance through "smart" devices and other networked sensors omnipresent in the Internet of Things. AI amplifies large-scale surveillance through techniques that analyze video, audio, images, and social media content across entire populations. The spread of smart robots in everyday life contributes to this development. As *Ryan Calo*[104] points out, robots not only greatly facilitate direct surveillance; they also introduce new points of access to historically protected spaces. Moreover, in becoming increasingly human-like, the social nature of robots may lead to new varieties of highly sensitive personal information.

In light of this development, there is growing doubt as to whether the existing data protection rules are sufficient to ensure adequate protection. This is particularly the case in countries such as the US, where data protection legislation is a patchwork of sector-specific laws that fail to adequately protect privacy.[105]

2. Frictions Between Big Data Practices Based on AI and the GDPR

The same cannot be said for the European Union. Since the General Data Protection Regulation (GDPR) came into force in May 2018, a high standard of personal data protection has been introduced in all member states – at least in theory.

However, there are increasing doubts as to whether the GDPR properly addresses the surge in Big Data practices and AI systems.

The GDPR applies to all personal data, meaning any information relating to an identified or identifiable natural person (Art. 4(1) GDPR). As most of the data that drives AI systems is either directly linked to a person, or, if anonymized, at least identifiable by an algorithm,[106] the GDPR applies regularly both when AI is under development (since it governs the collection and use of data in generating ML models) and also, under certain limited conditions, when it is used to analyze or reach decisions about

[104] *Calo*, Robots and Privacy, in: Lin/Abney/Beke (eds.), Robot Ethics: The Ethical and Social Implications of Robotics, 2011, pp. 187 et seq.

[105] According to *Solove*, Privacy and Power: Computer Databases and Metaphors for Information Privacy, (2001) 53 Stan. L. R. 1393, at p. 1430, the US system of data protection is one which "uses whatever is at hand (…) to deal with the emerging problems created by the information revolution."

[106] In the era of Big Data, anonymous information can be de-anonymized by employing related and non-related data about a person; *Barocas/Nissenbaum*, Big Data's End Run around Anonymity and Consent, in: Julia Lane et al. (eds.), Privacy, Big Data and the Public Good, 2014, pp. 49 et seq.; *Floridi*, The 4th Revolution, 2014, p. 110 *Rubinstein/Hartzog*, Anonymization and Risk, (2016) 91 Wash. L. Rev. 703, at pp. 710-711.

individuals. By contrast, there are no data protection rights or obligations concerning the ML models themselves in the period *after* they have been built but *before* any decisions have been taken about using them. As a rule, ML models do not contain any personal data, but only information about groups and classes of persons.[107] Although algorithmically designed group profiles may have a big impact on a person,[108] (ad hoc) groups are not recognized as holders of privacy rights. Hence, automated data processing by which individuals are clustered into groups or classes (based on their behavior, preferences, and other characteristics) creates a loophole in data protection law, pointing towards the need to recognize in the future some type of "group privacy" right.[109]

Beyond the issue of group privacy there is a series of further issues that show how little the GDPR takes into account the peculiarities of AI systems, self-learning algorithms, and Big Data Analytics, as many basic concepts and rules are in tension with these practices:[110]

- First of all, the *principle of purpose limitation* (Art. 5(1)(b) GDPR) is at odds with the prospect of Big Data analyses.[111] According to this principle, personal data must be collected for specified, explicit, and legitimate purposes and not further processed in a way incompatible with those purposes. However, analyzing Big Data quite often involves methods and usage patterns which neither the entity collecting the data nor the data subject considered or even imagined at the time of collection. Additionally, when it comes to ML algorithms it may be difficult to define the purpose of processing already at the stage of data collection because it is not possible to predict what the algorithm will learn. To inform the data subjects of the future forms of processing might prove costly, difficult, and even impossible.
- The *principle of data minimization* (Art. 5(1)(c) GDPR) also represents a challenging issue. Both Big Data and ML algorithms need a large amount of data to produce useful results. Arguably, the principle of data minimization does not mean that data controllers shall always collect as little data as possible, but only that the quantity must be related to the purpose provided that

[107] This could change due to evolving technologies. Cf. in particular *Veale/Binns/Edwards*, Algorithms that remember: model inversion attacks and data protection law, Phil. Trans. R. Soc. A 376: 20180083, http://dx.doi.org/10.1098/rsta.2018.0083, with the assumption that new forms of cyber attacks are able to reconstruct training data (or information about who was in the training set) in certain cases from the model.

[108] As *Hildebrandt*, Slaves to Big Data. Or Are We?, (2013) IDP Revista De Internet, Derecho y Política, pp. 27 et seq., at pp. 33 et seq., notes: "If three or four data points of a specific person match inferred data (a profile), which need not be personal data and thus fall outside the scope of data protection legislation, she may not get the job she wants, her insurance premium may go up, law enforcement may decide to start checking her email or she may not gain access to the education of her choosing."

[109] For further discussion, see: *Mittelstadt*, From Individual to Group Privacy in Biomedical Big Data, in: Cohen/Lynch/Vayena/Gasser (eds.), Big Data, Health Law, and Bioethics, 2018, pp. 175 et seq.; Taylor/Floridi/van der Sloot (eds), Group Privacy: New Challenges of Data Technologies, 1st ed., 2017.

[110] *Zarsky*, Incompatible: The GDPR in the Age of Big Data, (2017) 47(4) Seton Hall Law Review 995; *Humerick*, Taking AI Personally: How the E.U. Must Learn to Balance the Interests of Personal Data Privacy & Artificial Intelligence, (2018) 34 Santa Clara High Tech. L.J. 393. In contrast, the Information Commissioner's Office (ICO) in the UK does "not accept the idea that data protection, as currently embodied in legislation, does not work in a big data context", *ICO*, Big Data, Artificial Intelligence, Machine Learning, and Data Protection, 20170904, Version: 2.2 p. 95. Cf. also *Pagallo*, The Legal Challenges of Big Data: Putting Secondary Rules First in the Field of EU Data Protection, (2017) 3 European Data Protection Law Review 36, with reference to two possible solutions to make the collection and use of Big Data compatible with the GDPR: the use of pseudonymization techniques and the exemption of data processing for statistical purposes.

[111] *Forgó/Hänold/Schütze*, The Principle of Purpose Limitation and Big Data, in: Corrales/Fenwick/Forgó (eds.), New Technology, Big Data and the Law, 2017, pp. 17 et seq.

the data are adequate.[112] Nevertheless, this principle potentially undermines the utility and benefits of Big Data analyses.

- Third, it is problematic that the GDPR sets up a *special regime for particularly sensitive data*, e.g. data revealing not only racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, but also genetic data, biometric data and data concerning health, sex life, or sexual orientation (Art. 8 GDPR). Whereas the justification for setting a higher level of protection for special categories of data is intuitive, new forms of enhanced analytics challenge the ability to draw a clear distinction between "normal" personal data and "sensitive" data. After all, even an analysis merely relying on "regular" categories can quickly end up revealing sensitive data.
- Finally, AI-driven technologies also call into question another fundamental principle of data protection law, namely the *principle of consent*. How can data controllers possibly provide consent notices to individuals for potential secondary purposes that are yet to exist or have not been conceived? How can individuals have information regarding all of the possible implications communicated to them in comprehensible form, and be afforded the opportunity to understand what it is that they are being asked to consent to? How can algorithm-based profiling, nudging, and manipulation[113] be reconciled with freedom of choice and the idea of data protection as data subjects' control over their information?[114]

All these considerations show how little the new GDPR is compatible with big data analysis and AI products. Whether companies can comply with the requirements of the GDPR has yet to be proven. At the end of the day, much will depend on how the Regulation is interpreted by the courts and applied in practice. In this respect, two (extreme) scenarios are conceivable.[115] On the one hand, the GDPR might allow EU citizens to benefit from enhanced data protection, while still enjoying the innovations data analytics bring about.[116] On the other hand, the GDPR could threaten the development of AI, creating high market entry barriers for companies developing and/or using AI systems. According to this view, overregulation of personal data would lead to limited research and use of AI products. – Recent surveys show that such a scenario is not unlikely: Many companies see data protection as an obstacle to competition and are already complaining that AI products cannot be developed and distributed in the EU due to the strict rules.[117]

---

[112] *Noto La Diega*, Against the Dehumanisation of Decision-Making. Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information, (2018) 9(1) jipitec (Journal of Intellectual Property, Information Technology and E-Commerce Law) 1.

[113] Cf. *infra*, G.I.

[114] *Council of Europe*, Report on Artificial Intelligence. Artificial Intelligence and Data Protection: Challenges and Possible Remedies, report by *Alessandro Mantelero*, T-PD(2018)09Rev, https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6, p. 7. To address these issues, legal scholars have highlighted the potential role of transparency, or risk assessment as well as more flexible forms of consent, such as broad consent and dynamic consent; *Mantelero*, Regulating Big Data. The guidelines of the Council of Europe in the Context of the European Data Protection Framework (2017) 33(5) Computer Law & Sec. Rev. 584.

[115] *Zarsky* (n. 110).

[116] *Hildebrandt*, Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology, 2015, p. 211.

[117] Cf. *Delponte*, European Artificial Intelligence (AI) leadership, the path for an integrated vision, Study requested by the ITRE committee of the European Parliament, PE 626.074, September 2018, Figure 3 (Key barriers inhibiting faster deployment of AI systems in Europe), p. 17. According to surveys conducted by *Bitkom*, Germany's IT and telecommunications industry association, almost two-thirds of companies in Germany also say that data protection is an obstacle to the use of new technologies; (2018) Redaktion MMR-Aktuell, 406071.

For all these reasons, a thorough balancing seems necessary: If the EU wants to keep up with the global race to AI, it must carefully balance its interests in protecting personal data against its interest in developing new AI technologies.

III. Data Ownership vs. Data Access Rights

1. Protection of Data as (Intellectual) Property Rights?

Data has become the "new currency" in the digital world.[118] Data is collected by a variety of companies and converted into a valuable commercial product, which pays for many of the "free" services most consumers nowadays take for granted. Originally, Art. 3(3) of the proposal for an EU Digital Content Directive[119] explicitly mentioned the possibility of regarding personal data as a counter-performance (consideration) for the services received.[120] In B2B relationships, the possibility that (non-personal) data can be the subject of contractual agreements as commodities has been recognized even longer.[121] However, the problem with every "contractual approach" is that contractual obligations are only binding *inter partes*. Consequently, third parties cannot be prevented legally by contracts from using the data. In light of these considerations, there is an intensive discussion, especially in Europe, about whether a(n) (intellectual) property right in personal and/or non-personal data with *erga omnes* effect should be recognized.[122]

a) Personal Data

The discussion about possible property rights in data is not new. US scholars have been debating whether personal information should be viewed as property since the early 1970s.[123] The current debate, however, is based on very different premises. As *Purtova* points out, the propertization of personal information was viewed in the US mainly as an alternative to the existing data protection regime and one of the ways to fill in the gaps in the US data protection system.[124] This is different in Europe, where the GDPR provides a comprehensive set of data protection rules that in the end would interfere with the recognition of property rights in personal data: First of all, as the European Commission points out, such a property right would be incompatible with the fact that "the protection

---

[118] *Eggers/Hamill/Ali*, Data as currency, (2013) 13 Deloitte Review, p. 18 ff; *Langhanke/Schmidt-Kessel*, Consumer Data as Consideration, (2018) Journal of European Consumer and Market Law (EuCML) 218; *Taylor*, Data: the new currency, 2014. The *European Commission*, Communication "Building a European Data Economy", COM(2017) 9 final, predicts that the value of the European data economy will increase to EUR 643 billion by 2020, representing 3.17% of the overall EU GDP.

[119] Art. 3(3) of the Proposal for a Directive on certain aspects concerning contracts for the supply of digital content, COM(2015) 634 final, stated that the Directive "shall apply to any contract where the supplier supplies digital content to the consumer or undertakes to do so and, in exchange, a price is to be paid or the consumer actively provides counter-performance other than money in the form of personal data or any other data".

[120] By contrast, Art. 3(1) of the Digital Content Directive no longer uses the term "counter-performance" in order to mitigate the concerns about treating personal data as a commodity; cf. the concerns of the European Data Protection Supervisor, Opinion 4/2017 on the Proposal for a Directive on certain aspects concerning contracts for the supply of digital content, March 14, 2017, pp. 7-9 and pp. 16-17.

[121] COM(2017) 228 final, under 3.2.; SWD(2017) 2 final, p. 16; cf. *Berger*, Property Rights to Personal Data? – An Exploration of Commercial Data Law, (2017) Zeitschrift für geistiges Eigentum (ZGE) 340: „data contract law lies at the heart of commercial data law."

[122] For an overview of the academic discussion in several countries cf. *Osborne Clarke LLP*, Legal study on Ownership and Access to Data, Study prepared for the European Commission DG Communications Networks, Content & Technology, 2016.

[123] *Westin*, Privacy and Freedom, 1967; *Solove* (n. 105), at pp. 1421-1422; *Lessig*, Privacy as Property, (2002) 69(1) Social Research: An International Quarterly of Social Sciences 247; *Schwarz*, Property, privacy and personal data, (2004) 117(7) Harvard L Rev 2055.

[124] *Purtova*, Property rights in personal data: Learning from the American discourse, (2009) Computer Law & Security Review 507.

of personal data enjoys the status of a fundamental right in the EU".[125] In addition, a property right in personal data would be inconsistent with Art. 7(3) GDPR according to which consent can be withdrawn even against the will of the entitled legal entity. Finally, even if a right to one's data was constituted, it would remain a challenge to assign such a right to one single person, as most personal data relates to more than one data subject.[126]

b) Non-personal Data

Admittedly, these problems do not exist with non-personal data ("pure" machine-generated data). As non-personal data is neither protected by data protection law nor *as such* by (European) IP law,[127] some scholars recently argued in favor of the creation of a new property right with the objective of enhancing the tradability of anonymized machine-generated data.[128] The European Commission also temporarily considered the introduction of a "data producer's right" with the aim of "clarifying the legal situation and giving more choice to the data producer, by opening up the possibility for users to utilize their data".[129]

Still, there are serious concerns about the introduction of such a right: Firstly, there is no practical need for such a property right, since companies can effectively control the access to "their" data by technical means. Secondly, companies "possessing" data are protected through a number of other legal instruments (e.g. tort and criminal law) against destruction, certain impediments to access and use, as well as against compromising their integrity.[130] Thirdly, the legal discussion has shown that the specification of the subject matter and the scope of protection seems to be extremely difficult in regard to data.[131] Last but not least, the introduction of an exclusive right to data bears the serious risk of an inappropriate monopolization of data.[132] Granting data holders an absolute (intellectual) property right on data would strengthen their (dominant) position, increasing entry barriers for competitors.

It is therefore fitting that the European Commission no longer appears to be pursuing the discussion on the introduction of data ownership rights and is instead concentrating on the question of how to deal with data-driven barriers to entry.

---

[125] *European Commission*, Staff Working Document on the free flow of data and emerging issues of the European data economy accompanying the document Communication Building a European data economy, 10.01.2017, SWD(2017) 2 final, p. 24.

[126] *Purtova*, Do property rights in personal data make sense after the big data turn: Individual control and transparency, (2017) 10(2) Journal of Law and Economic Regulation 64.

[127] Raw machine-generated data are not protected by existing IP rights since they are not deemed to be the result of an intellectual effort and/or have no degree of originality. Likewise, the Database Directive 96/9/EC does not protect data as such, but only data originating from a protected database. Similarly, the Trade Secrets Directive 2016/943, does not grant an absolute right to data but is based on the maintenance of factual secrecy; as *Wiebe*, Protection of industrial data – a new property right for the digital economy?, (2016) Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil (GRUR Int.) 877, points out: "Once secrecy is lost, legal protection is lost as well".

[128] Cf. in particular *Zech*, Data as a Tradeable Commodity, in: de Franceschi (ed.), European Contract Law and the Digital Single Market. The Implications of the Digital Revolution, 2016, pp. 51 et seq.; *Becker*, Rights in Data. Industry 4.0 and the IP-Rights of the Future, (2017) 9 ZGE/Intellectual Property Journal (IPJ) 253.

[129] *European Commission*, Communication "Building a European Data Economy, COM(2017) 2 final, p. 13; cf. moreover Commission Staff Working Document (n. 125), pp. 33 et seq.

[130] *Kerber*, A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis, (2016) GRUR Int. 989.

[131] *Wiebe* (n. 127), at pp. 881-883.

[132] *Max Planck Institute for Innovation and Competition*, Position Statement of 26 April 2017 on the European Commission's "Public consultation on Building the European Data Economy, p. 6; *Drexl*, Neue Regeln für die Europäische Datenwirtschaft? Ein Plädoyer für einen wettbewerbspolitischen Ansatz – Teil 1, (2017) Neue Zeitschrift für Kartellrecht (NZKart) 339, at p. 343.

2. Access to Data

The European Commission acknowledges a growing concern that the control of large volumes of data could lead to situations of market power.[133] In the same vein, the OECD points out that larger incumbents – due to the network effects previously discussed[134] – are likely to benefit from significant advantages over smaller firms and "second movers" in collecting, storing, and analyzing large and heterogeneous types of data.[135] Smaller firms and new entrants might therefore face barriers to entry, preventing them from developing algorithms that can effectively exert competitive pressure.

Some argue we only need to apply competition law and split up internet giants, like Standard Oil or AT&T in decades past.[136] Others believe that the appropriate remedy against a concentration of data in the hands of too few is aggressive antitrust action and a mandate for companies to share proprietary data proportional to market share. In this spirit, *Mayer-Schönberger/Ramges* propose in their book "Reinventing Capitalism" a progressive data-sharing mandate which would require Facebook (and any similarly structured powerful player) to share proprietary data proportional to market share.[137] – However, both demands can hardly be realized on the basis of current competition law. According to many legal systems, an unbundling of an entire company is only permissible – if at all – in cases where it repeatedly violates competition law in a particularly serious manner.[138] The *essential facility doctrine*, under which a company with a dominant position must grant access to a facility under specific conditions,[139] does not help either, because this doctrine only applies under "extraordinary circumstances".[140]

---

[133] EU Commissioner *Vestager*, Competition in a big data world. Paper presented at the Digital Life Design (DLD) Conference, 2016, available at https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/competition-big-data-world_en. Cf. moreover *Rubinfeld/S. Gal*, Access Barriers to Big Data, 2017 (59) Arizona Law Review 339; *Vezzoso*, Competition policy in a world of big data, in: Olleros/Zhegu (eds.), Research Handbook on Digital Transformations, 2016, pp. 400 et seq.

[134] Cf. *supra*, F.I.

[135] *OECD*, Big Data: Bringing Competition Policy to the Digital Era, 2016, https://one.oecd.org/document/DAF/COMP(2016)14/en/pdf.

[136] In this sense, for example *Galloway*, Silicon Valley's Tax-Avoiding, Job-Killing, Soul-Sucking Machine, Esquire, March 2018, https://www.esquire.com/news-politics/a15895746/bust-big-tech-silicon-valley/?src=nl&mag=esq&list=nl_enl_news&date=020818.

[137] *Mayer-Schönberger/Ramges* (Fn. 102).

[138] For the EU, cf. Regulation 1/2003, recital (12): "Changes to the structure of an undertaking as it existed before the infringement was committed would only be proportionate where there is a substantial risk of a lasting or repeated infringement that derives from the very structure of the undertaking." For the USA, cf. Sec 2 of the Sherman Antitrust Act 1890: "Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty (…)".

[139] For the US, see MCI Commc'ns Corp. v. American Tel. & Tel. Co., 708 F.2d 1081, 1132–33 (7th Cir. 1983); *Maurer/Scotchmer*, The Essential Facilities Doctrine: The Lost Message of Terminal Railroad, March 10, 2014, UC Berkeley Public Law Research Paper No. 2407071, https://ssrn.com/abstract=2407071; *Pitofsky/Patterson/Hooks*, The Essential Facilities Doctrine Under US Antitrust Law, (2002) 70 Antitrust Law Journal 443, at p. 448. For the EU, see ECJ, 6.5.1995, joined cases C-241-242/91 P (*RTE and ITP/Kommission* – „*Magill*"), ECLI:EU:C:1995:98; 29.4.2004, case C-418/01 (*IMS Health*), ECLI:EU:C:2004:257; CFI, 17.9.2007, case T-201/04 (*Microsoft/Commission*), ECLI:EU:T:2007:289; *Evrard*, Essential Facilities in the European Union: Bronner and Beyond, (2004) 10 Columbia Journal of European Law 491.

[140] On the question of whether data can be regarded as an essential facility, cf. from a US American perspective *Sokol/Comerford*, Antitrust and Regulating Big Data, (2016) 23 Geo. Mason L. Rev., 1129, at pp. 1158 et seq.; *Balto,* Monopolizing Water in a Tsunami: Finding Sensible Antitrust Rules for Big Data, 2016, http://ssrn.com/abstract=2753249. For the European perspective cf. *Graef*, Data as Essential Facility. Competition and Innovation on Online Platforms, PhD Theses, KU Leuven, 2016, *https://core.ac.uk/download/pdf/34662689.pdf*; *Lehtioksa*, Big Data as an Essential Facility: the Possible Implications for Data Privacy, Master's thesis, University of Helsinki, 2018, https://www.paulo.fi/sites/default/files/inline-files/Lehtioksa%20Jere_pro%20gradu.pdf; *Telle*, Kartellrechtlicher Zugangsanspruch zu Daten nach der essential facility doctrine, in: Hennemann/Sattler (eds.), Immaterialgüter und Digitalisierung, 2017, pp. 73-87.

Apart from this, antitrust law is a very limited tool for mandating access to data, mainly for three reasons. First, in dynamic multi-sided markets it is very difficult to prove the existence of a monopolistic position and/or market dominance[141] and establish clear criteria for exploitative abuse in regard to data. Secondly, competition law is generally unable to limit the price that can be set by the data monopolist in exchange for access. And third, antitrust law does not deal effectively with situations in which market power arises from oligopolistic coordination.[142]

For all these reasons, it seems more promising to create specific statutory data access rights. In the European Union, such rights already exist in specific contexts.[143] Accordingly, there are models upon which the European legislature could build. A general right of access to data applicable to all sectors, on the other hand, does not seem appropriate. Rather, a targeted approach is to be preferred[144] which, depending on the sector, attempts to balance the legitimate interest of persons in access to external data with the legitimate interest of data generators (or data holders) in the protection of their investments and – where personal data is involved – the interests of data subjects.

G. Algorithmic Manipulation and Discrimination of Citizens, Consumers, and Markets

Self-learning algorithms are used by many companies, political parties, and other actors to influence and manipulate citizens and consumers through microtargeting. This raises the question of how the law can provide adequate safeguards against such practices (I.). Another problem closely related to algorithmic decision making is the risk of discrimination: Many studies indicate that algorithms are often not value neutral, but biased and discriminatory. Here, too, the question arises as to what extent citizens and consumers can and should be protected (II.). Beyond these issues, the phenomenon of algorithmic manipulation and discrimination also poses interesting competition law questions in cases where algorithms interact collusively (III.).

I. Profiling, Targeting, Nudging, and Manipulation of Citizens and Consumers

1. The Technique of Behavioral Microtargeting

In recent years, behavioral microtargeting has developed into a new, promising business strategy. The technique of behavioral microtargeting allows companies to address people individually according to their profile, which is created algorithmically from personal data about the individual's behavior and personality.[145]

By and large, behavioral microtargeting is based on three elements. The psychometric analysis of individuals requires *first* the collection of large amounts of data. In a *second* step, the collected data is evaluated by machine learning algorithms in order to analyze and predict certain personal traits of users: their character strengths, but also their cognitive and voluntative weaknesses. In this regard, several studies by researchers from the University of Cambridge have shown that the analysis of

[141] Traditional approaches to market definition fail with digital platforms because (i) many platforms work with free goods and services and (ii) are characterized by having several market sides, which makes it very difficult to assess the competitive powers at play; cf. *Podszun/Kreifels*, Digital Platforms and Competition Law, (2016) EuCML 33.

[142] *OECD*, Directorate for Financial and Enterprise Affairs Competition Committee, Competition Enforcement in Oligopolistic Markets – Issues paper by the Secretariat, 16-18 June 2015, DAF/COMP(2015)2.

[143] Cf. for example Art. 6-9 Regulation 715/2007/EC, Art. 35-36 Directive 2015/2366/EU, Art. 27, 30 Regulation 2006/1907/EC, Art. 30, 32 Directive 2009/72/EC and Recital 11 Directive 2010/40/EU. The right to portability embodied in Art. 20 GDPR is also based on the ratio to avoid lock-in effects and to improve the switching process from one service provider to another.

[144] Similarly, *Max Planck Institute for Innovation and Competition*, Position Statement of 26 April 2017 on the European Commission's "Public consultation on Building the European Data Economy", p. 11.

[145] *Calo*, Digital Market Manipulation, (2014) 82(4) The George Washington Law Review 995, at pp. 1015 et seq.; *O'Neil*, Weapons of Math Destruction, 2016, pp. 194 et seq.; *European Data Protection Supervisor (EDPS)*, Opinion 3/2018 on online manipulation and personal data, March 19, 2018.

(neutral) Facebook "likes" provides far-reaching conclusions about the personality of an individual.[146] According to these studies, an average of 68 Facebook "likes" suffices to determine the users' skin color (with 95 % accuracy), sexual orientation (88 % accuracy), and affiliation to the Democratic or Republican party (95 % accuracy). In addition, the studies claim that it is possible to use Facebook "likes" to predict religious affiliation; alcohol, cigarette, and drug consumption; as well as whether or not a person's parents stayed together until that person reached the age of 21. With the input of even more Facebook "likes," the algorithm was able to evaluate a person better than the person's friends, parents, and partners, and could even surpass what the person thought they knew about themselves.[147]

The processed data can be used, in a *third* step, in a variety of ways. Companies can tailor their advertising campaigns but also their products and prices specifically to the customer profile,[148] credit institutions can use the profiles for credit rating,[149] insurance companies can better assess the insured risk,[150] HR departments can pre-select candidates,[151] and parties can use the data for political campaigns – a practice which in the end led to the well known Cambridge Analytica scandal.[152] In the US, the judiciary system is now using big data analysis to predict the future behavior of criminals.[153]

2. Behavioral Economics and Behavioral Microtargeting

Combining big data with findings in behavioral economics leads to some noteworthy insights on microtargeting: For some time now, economists have been shifting away from the paradigm of economic neoclassicism, the *homo oeconomicus*, whose guiding principle is based on the assumption that individuals make rational decisions.

By contrast, behavioral economics has been able to show that humans have only limited rationality, primarily because of cognitive limitations of the human mind (bounded rationality), but also because humans often take actions that they know to be in conflict with their own long-term interests (bounded willpower), and, moreover, their care about others (bounded self-interest).

Modern market research tries to exploit these "vulnerabilities" and combines them with Big Data. In this respect, mounting empirical evidence shows that companies are exploiting or even trying to cause irrational behavior:

- In 2014, Facebook manipulated the newsfeeds of over half a million users in order to alter the emotional content of users' posts, showing in this experiment that user feelings can be deliberately manipulated by certain messages (so-called emotional contagion).[154]

---

[146] *Kosinski/Stillwell/Graepel*, (2013) 110(15) PNAS 5802, http://www.pnas.org/content/110/15/5802.full; *Youyou/Kosinski/Stillwell*, Computer-based personality judgments are more accurate than those made by humans, (2015) 112(4) PNAS 1036, http://www.pnas.org/content/112/4/1036.full.

[147] Summarizing *Grassegger/Krogerus*, The Data That Turned the World Upside Down, Motherboard, January 28, 2017, https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win.

[148] *Hofmann*, Der maßgeschneiderte Preis, (2016) Wettbewerb in Recht und Praxis (WRP) 1074; *Zuiderveen Borgesius/Poort*, Online Price Discrimination and EU Data Privacy Law, (2017) 40 J Consum Policy 347.

[149] Cf. *Citron/Pasquale*, (2014) 89 Washington Law Review 1; *Zarsky*, Understanding Discrimination in the Scored Society, (2014) 89 Washington Law Review 1375.

[150] Cf. *Swedloff*, Risk Classification's Big Data (R)evolution, (2014) 21.1 Connecticut Insurance Law Journal 339; *Helveston*, Consumer Protection in the Age of Big Data, (2016) 93(4) Washington University Law Review 859.

[151] Cf. *O'Neil* (n. 145), pp. 105 et seq.

[152] Cf. the speech by *Alexander Nix*, ex CEO of Cambridge Analytica, at the 2016 Concordia Annual Summit in New York, https://www.youtube.com/watch?v=n8Dd5aVXLCc; moreover *Rubinstein*, Voter Privacy in the Age of Big Data, (2014) Wisconsin Law Review 861; *Hoffmann-Riehm*, (2017) 142 Verhaltenssteuerung durch Algorithmen, Archiv des öffentlichen Rechts (AöR) 1.

[153] *Angwin* et al., Machine Bias, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[154] *Goel*, Facebook tinkers with users' emotions in news feed experiment, stirring outcry, New York Times, June 29, 2014, https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feedexperiment-stirring-outcry.html?_r=0; *Kramer/Guillory/Hancock*, Experimental evidence of massive-scale

- In early 2017, it also became known that Facebook Australia had offered its advertisers a software that could accurately locate psychologically unstable, depressed teenagers.[155]
- In 2012, Microsoft registered a patent on "Targeting Advertisements Based on Emotion".[156] And in 2013, Samsung filed the patent "Apparatus and methods for sharing user's emotion".[157]

3. Algorithmic Echo Chambers, Filter Bubbles, and Fake News: A Danger to Democracy?

The use of algorithms to channel information on social media platforms and search engines has led to a growing fear that the use of content-filtering and content-removing AI systems as well as social media bots spreading political messages will have a detrimental effect on the right to freedom of information, the right to freedom of expression, media pluralism, and political discourse in general. Following the US elections in 2016, public concern has also grown with respect to the creation and dissemination of fake news and its influence over democratic decision-making processes.

Indeed, algorithm-based search engines and social networks can channel and control a variety of factors that affect how opinions are formed. In many cases, algorithms (and social bots) determine which content is selected, processed, and published; sometimes algorithms and social bots are even used to create new content. The "master" of the algorithm is thus to a large extent also the "ruler" of public opinion: Whoever configures the respective algorithm makes essential decisions regarding the information displayed and thus influences opinions.

The use of algorithms combined with the increasing monopolization of market power and knowledge in the platform economy[158] can lead in particular – so it is feared – to so-called "echo chambers", in which people encounter only information that confirms their existing political views.[159] A related theory about "filter bubbles" claims that algorithms cause bubbles of like-minded content around news users.[160] For these reasons, there are serious concerns both in the US and in Europe that (media) diversity could be drastically reduced.[161] Moreover, AI systems create new opportunities to enhance "fake news" by simplifying the production of high-quality fake video footage; automating the writing and publication of fake news stories; and microtargeting citizens, delivering the right message at the right time in order to maximize persuasive potential.[162]

In light of these considerations, a number of (regulatory) issues are discussed:[163] Are information intermediaries such as Facebook and Google simply hosts of user-created content, or have they already turned into media companies themselves? At which point is it no longer justified to maintain the differences in (self) regulation between the traditional media and these platforms in terms of

---

emotional contagion through social networks, (2014) 111(24) PNAS 8788, www.pnas.org/content/111/24/8788.full.pdf.

[155] *Davidson*, Facebook targets 'insecure' young people, The Australian, May 1, 2017; cf. also http://www.news.com.au/technology/online/social/leaked-document-reveals-facebook-conducted-research-to-target-emotionally-vulnerable-and-insecure-youth/news-story/d256f850be6b1c8a21aec6e32dae16fd.

[156] Microsoft Corporation (2012): Targeting Advertisements Based on Emotion, US 20120143693 A1, http://www.google.com/patents/US20120143693.

[157] Samsung Electronics Co., Ltd. (2013): Apparatus and method for sharing user's emotion. US 20130144937 A1, http://www.google.com/patents/US20130144937.

[158] Cf. *supra*, F.I. and F.III.2.

[159] *Cass R. Sunstein*, #Republic. Divided Democracy in the Age of Social Media, 2017.

[160] *Pariser*, The Filter Bubble: What the Internet is Hiding from You, 2011.

[161] *Epstein*, How Google Could End Democracy, U.S. News & World Report, June 9, 2014, https://www.usnews.com/opinion/articles/2014/06/09/how-googles-search-rankings-could-manipulate-elections-and-end-democracy. See also the 2016 Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, to the 32nd session of the Human Rights Council (A/HRC/32/38), noting that "search engine algorithms dictate what users see and in what priority, and they may be manipulated to restrict or prioritise content".

[162] *Brundage* et al (n. 84), pp. 43 et seq.

[163] *Helberger/Kleinen-von Königslöw/van der Noll*, Regulating the new information intermediaries as gatekeepers of information diversity, (2015) 17(6) Info 50, http://www.ivir.nl/publicaties/download/1618.pdf.

advertising regulation, taxation, program standards, diversity, and editorial independence? What are the responsibilities of information intermediaries regarding fake news and filtering information in general? Should users be (better) informed about the personalization of (news) content? Do we want to limit the personalization of information/communication by legislation? Is it perhaps even necessary to regulate the algorithm itself in order to ensure adequate diversity of media and opinion?
Although these questions certainly need to be addressed, it should also be noted that there is still no established scientific evidence for the existence of echo chambers and filter bubbles. Recently published studies claim that these fears might be blown out of proportion, because most people already have media habits that help them avoid echo chambers and filter bubbles.[164] Moreover, it is unclear to what extent political bots spreading fake news succeed in shaping public opinion, especially as people become more aware of these bots' existence.[165] In this light, the call for legislation appears premature. What is needed above all are further empirical studies examining the effect of algorithm-driven information intermediaries more closely.

4. Manipulation of Consumers: The Case of Exploitative Contracts

The use of microtargeting techniques also leads to new forms of information asymmetries between contractual partners, and to an erosion of private autonomy.[166] AI driven big data profiling techniques give companies the opportunity to gain superior knowledge about customers' personal circumstances; behavioral patterns; and personality, including future preferences. These insights enable companies to tailor contracts in ways that maximize their expected utility by exploiting the behavioral vulnerabilities of their clients. Behavioral economics has identified hundreds of effects, all of which demonstrate that human decision-making behavior is irrational in many situations, but nevertheless predictable and can be exploited accordingly. Microtargeting makes it possible, for instance, to offer products exactly when the customer can only make sub-optimal decisions – for example, due to the time of day or a previous event. This so-called *emotional targeting* is already being used by many companies. For example, the US advertising company MediaBrix developed a system that analyzes the emotions of computer players in real time and then addresses them directly through personalized advertising at particularly suitable moments (during breakthrough moments).[167]

This example alone demonstrates that behavioral microtargeting has a high potential for abuse: based on the findings of behavioral economics, companies can exploit or even induce suboptimal decision-making behaviors in their customers.

Existing European consumer and data protection law as well as national contract law arguably fail to provide sufficient instruments to effectively sanction such behavior.

---

[164] *Dubois/Blank*, The echo chamber is overstated: the moderating effect of political interest and diverse media, (2018) 21(5) Information, Communication & Society 1; *Moeller/Helberger*, Beyond the filter bubble: concepts, myths, evidence and issues for future debates, June 25, 2018, http://hdl.handle.net/11245.1/478edb9e-8296-4a84-9631-c7360d593610.

[165] *Nyhan*, Fake News and Bots May Be Worrisome, but Their Political Power Is Overblown, The New York Times, February 13, 2018, https://www.nytimes.com/2018/02/13/upshot/fake-news-and-bots-may-be-worrisome-but-their-political-power-is-overblown.html; *Brundage* et al. (n. 84), p. 46. Cf. also *Kalla/Broockman*, The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments, September 25, 2017, forthcoming, American Political Science Review; Stanford University Graduate School of Business Research Paper No. 17-65, https://ssrn.com/abstract=3042867.

[166] *Mik*, The Erosion of Autonomy in Online Consumer Transactions, (2016) 8(1) Law, Innovation and Technology 1, http://ink.library.smu.edu.sg/sol_research/1736; *Sachverständigenrat für Verbraucherfragen (SVRV)*, Verbraucherrecht 2.0, Verbraucher in der digitalen Welt, December 2016, pp. 58 et seq., http://www.svr-verbraucherfragen.de/wp-content/uploads/Gutachten_SVRV-.pdf.

[167] *Pritz*, Mood Tracking: Zur digitalen Selbstvermessung der Gefühle, in: Selke (ed.), Lifelogging, 2016, pp. 127 et seq., at pp. 140 et seq.

First of all, it is questionable whether microtargeting can be classified as an unfair commercial practice according to the Unfair Commercial Practices Directive (UCPD). As *Eliza Mik*[168] and others[169] have pointed out, the main weaknesses of the UCPD lie in the definitions and assumptions underlying the concepts of "average" and "vulnerable" consumers (which disregard the findings in behavioral economics and cognitive science) as well as the narrow definition of aggressive practices such as undue influence, which requires the presence of pressure, thus failing to address cases of subtler forms of manipulation. A similar picture emerges for European data protection law, which suffers – above all – from an overreliance on control and rational choice that vulnerable users are unlikely to exert.[170]

Whether these gaps in protection can be compensated by (national) contract law is also questionable since it is difficult to subsume microtargeting under any of the traditional protective doctrines – such as duress, mistake, undue influence, misrepresentation, or culpa in contrahendo.[171] At the end of the day, the impact of microtargeting on customer behavior appears to be too subtle to be covered by common concepts of contract law, despite the fact that such a technique affects one of its central values: autonomy.

Future regulation will therefore have to evaluate the extent to which customers should be protected from targeted advertisements and offers that seek to exploit their vulnerabilities. This is by no means an easy task because – as *Natali Helberger*[172] rightly points out – there is a very fine line between informing, nudging, and outright manipulation.

II. Discrimination of Citizens and Consumers

1. How AI Systems Can Lead To Discrimination

The widespread use of algorithms for preparing or even making decisions, some of which may have existential significance for people, is increasingly criticized by policymakers around the world on the grounds of discrimination.[173] In fact, a number of examples show that ADM procedures are by no means neutral, but can perpetuate and even exacerbate human bias in various ways.
Examples include a chatbot used by Microsoft who unexpectedly learned how to post racist and sexist tweets,[174] a face recognition software used by Google which inadvertently classified black people as

[168] *Mik (*n. 166).
[169] *Ebers*, Beeinflussung und Manipulation von Kunden durch "Behavioral Microtargeting", (2018) MultiMedia und Recht (MMR) 423; *Duivenvoorde*, The Protection of Vulnerable Consumers under the Unfair Commercial Practices Directive, (2013) 2 Journal of European Consumer and Market Law 69.
[170] *Hacker*, Personal Data, Exploitative Contracts, and Algorithmic Fairness: Autonomous Vehicles Meet the Internet of Things, (2017) 7 International Data Privacy Law 266, https://ssrn.com/abstract=3007780.
[171] Cf. *Mik*, (n. 166).
[172] *Helberger*, Profiling and Targeting Consumers in the Internet of Things – A New Challenge for Consumer Law, in: Schulze/Staudenmayer (eds.), Digital Revolution: Challenges for Contract Law in Practice, 2016, pp. 135 et seq., at p. 152.
[173] *Executive Office of the [US] President*, Preparing for the Future of Artificial Intelligence (Report, 2016) 30-32; *Powles*, New York City's Bold, Flawed Attempt to Make Algorithms Accountable, New Yorker (December 20, 2017), https://www.newyorker.com/tech/elements/new-york-citys-bold-flawed-attempt-to-make-algorithmsaccountable; *European Parliament*, Resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (March, 2017), Art. 19-22; and for Germany: Wissenschaftliche Dienste des Deutschen Bundestags, Einsatz und Einfluss von Algorithmen auf das digitale Leben, Aktueller Begriff (October 27, 2017).
[174] See *Vincent*, Twitter Taught Microsoft's Friendly AI Chatbot to Be a Racist Asshole in Less than a Day, The Verge, March 24, 2016, https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

gorillas,[175] and the COMPAS algorithm which is increasingly used by US courts to predict the likelihood of recidivism of offenders: As the news portal ProPublica revealed in 2016, COMPAS judged black and white prisoners differently. Among other things, it was found that the probability that black inmates were identified as high-risk but did not re-offend, was twice as high as for white inmates. Conversely, white inmates were more likely to be classified as low-risk, but later to re-offend.[176]
Such discrimination can have various causes.[177]
Discrimination occurs primarily at the *process level*[178] when the algorithmic model is fed with biased training data. Such bias can take two forms.[179] One occurs when errors in data collection lead to inaccurate depictions of reality due to improper measurement methodologies, especially when conclusions are drawn from incorrect, partial, or nonrepresentative data. This type of bias can be addressed by "cleaning the data" or improving the data collection process. The second type of bias occurs when the underlying process draws on information that is inextricably linked to structural discrimination, exhibiting long-standing inequality. This happens, for example, when data on a job promotion is collected from an industry in which men are systematically favored over women. In this scenario, the data basis itself is correct. However, by using this kind of data in order to decide whether employees are worthy of promotion, a discriminatory practice would be perpetuated and continued in the future.
Apart from biased training data, discrimination can also be caused at the *classification level*[180] by feature selection, for example by using certain protected characteristics (such as race, gender, or sexual orientation) or by relying on factors that happen to serve as proxies for protected characteristics (e.g. using the place of residence in areas that are highly segregated).[181]

2. Anti-Discrimination Law

Although there is extensive anti-discrimination legislation in both the US and the European Union, the problem of algorithmic discrimination is insufficiently addressed on both sides of the Atlantic. In the US, this is partly due to the fact that anti-discrimination legislation is limited primarily to the employment sector.[182] Besides, there are a number of other reasons why discriminatory algorithmic systems often escape the doctrinal categories of US anti-discrimination law, or, more precisely, Title VII of the Civil Rights Act of 1964. As *Barocas & Selbst* have highlighted, this is mainly the case because (i) the disparate treatment doctrine focuses on human decision makers as discriminators without taking into account unintentional discrimination, and (ii) decision makers can often escape disparate impact liability if the factors used for data-mining are job-related.
Likewise, EU anti-discrimination law does not provide adequate protection against algorithmic discrimination.[183]

---

[175] *Barr*, Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms, Wall Street Journal, July 1, 2015, https://blogs.wsj.com/digits/2015/07/01/google-mistakenly-tags-black-people-as-gorillasshowing-limits-of-algorithms/.
[176] See *Larson* et al., How We Analyzed the COMPAS Recidivism Algorithm, ProPublica, May 23, 2016, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm; *Kleinberg* et al., Inherent trade-offs in the fair determination of risk scores, Working Paper (2016), https://arxiv.org/abs/1609.05807, at pp. 5-6.
[177] See for example, *Barocas/Selbst*, Big Data's Disparate Impact, (2016) 104 California Law Review 671, at p. 680; *Kroll* et al., Accountable Algorithms, (2016) 165 University of Pennsylvania Law Review 633, at pp. 680 et seq.
[178] For the different dimensions (process, model, and classification level) cf. *supra*, B.IV.
[179] *Crawford/Whittaker*, The AI Now Report, The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term, 2016.
[180] Cf. again *supra*, B.IV.
[181] *Kroll* et al. (177), at pp. 681 et seq.
[182] For a comparison between US and EU anti-discrimination law cf. *de Búrca*, The Trajectories of European and American Antidiscrimination law, (2012) 60 American Journal of Comparative Law 1.
[183] Cf. *Hacker*, Teaching Fairness To Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination under EU Law, (2018) 55 Common Market Law Review (CMLR) 1143.

Problems arise, first of all, with regard to the limited scope of EU anti-discrimination Directives. Although the Race Equality Directive 2000/43/EC and the Gender Equality Directive 2004/113/EC extend equal treatment principles beyond employment matters far into general contract law, their scope is nevertheless limited, because they only apply (i) to race and gender discrimination and (ii) when goods or services are "available to the public".[184] Both limitations appear to be problematic: On the one hand, the respective Directives do not cover other discriminatory factors such as religion or belief, disability, age, sexual orientation, or financial status and willingness to pay,[185] nor (new) types of AI-driven differentiations which treat people unequally because they belong to a specific group (as for example the group of cat lovers or Nike shoe wearers).[186] On the other hand, there is the problem that, due to the use of microtargeting, offers and contracts are increasingly tailored and personalized, which raises the question of whether such goods or services are any longer "available to the public".[187] Moreover, anti-discrimination law does not address the possibility that a prediction may prove to be wrong in a particular case. If, for example, the predictive model is based on the assumption that 80 % of the people living in a certain area pay their bills late, and a company denies loans to all people living there, it also denies loans to the 20 % who pay their bills on time.[188] In this case, too, the outcome of the assessment is of course unfair. Such a result is, however, not due to a discriminatory practice, but to the fact that statistic models do not consider individual cases but rather generalize them. In these scenarios, the tricky question is what degree of individual fairness is required and how much generalization can be accepted.

Finally, many biased decisions which amount to indirect discrimination can be justified if the predictive task of the ADM process furnishes a legitimate aim (such as future job performance, credit worthiness, etc.).[189] In these cases, the victim has to "prove the model wrong" by establishing, for example, that the seemingly high predictive value of the AI system stems from biased training data. Doing so is no easy task, however, as victims of algorithmic discrimination will be unable to establish even a prima facie case of discrimination without access to the data and algorithms, and in many cases do not even know they have been the victim of discrimination at all

3. Discussion

In view of this situation, various solutions are being discussed for both the US and the European Union. With regard to individual enforcement, the following measures are proposed in particular: (i) information rights regarding the scoring process; (ii) duties to provide consumers with tools for

---

[184] Art. 3(1)(h) Race Equality Directive 2000/43/EC; Art. 3(1) Gender Equality Directive 2004/113.

[185] On the problem of so-called first-degree price-discrimination, see *European Data Protection Supervisor*, (2015), Opinion No. 7/2015, Meeting the challenges of big data: A call for transparency, user control, data protection by design and accountability, https://edps.europa.eu/data-protection/our-work/publications/opinions/meeting-challenges-big-data_en; *Article 29 Working Party*, Opinion 03/2013 on purpose limitation, http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf; *Bar-Gill*, Algorithmic Price Discrimination: When Demand is a Function of Both Preferences and (Mis)perceptions, May 29, 2018, The Harvard John M. Olin Discussion Paper Series, No. 05/2018; Harvard Public Law Working Paper No. 18-32, https://www.ssrn.com/abstract=3184533; *Zuiderveen Borgesius/Poort* (n. 148). EU competition law prohibits different prices only if a company abuses its dominant position; cf. esp. Art 102(2)(a) and (c) TFEU. EU consumer protection rules, in particular the Unfair Commercial Practices Directive 2005/29/EC, also leave traders free to set prices as long as they inform consumers about the prices and how they are calculated; *European Commission*, Guidance on the Implementation/Application of Directive 2005/29/EC on Unfair Commercial Practices, SWD(2016) 163 final, p. 134.

[186] *Martini*, in this book, Chapter 3; *Zuiderveen Borgesius*, Discrimination, artificial intelligence, and algorithmic decision-making, Study for the Council of Europe, 2018, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73, pp. 35 et seq.

[187] *Hacker* (n. 183), at pp. 1156 et seq.; *Busch*, Algorithmic Accountability, ABIDA Project Report, March 2018, http://www.abida.de/sites/default/files/ABIDA%20Gutachten%20Algorithmic%20Accountability.pdf, p. 47.

[188] *Zuiderveen Borgesius* (n. 186), p. 36.

[189] *Hacker* (n. 183), at pp. 1160 et seq.; *Zuiderveen Borgesius* (n. 186), pp. 19 et seq.

interactive modeling; (iii) access rights to data sets; or, alternatively, (iv) a right to confidential review of the logics of predictive scoring, including the source code (e.g. by trusted third parties) in order to challenge decisions based on ADM procedures. In the EU, it is disputed above all whether a right to explanation of automated decision making can be derived already from the GDPR itself.[190]

In addition to individual remedies, a number of other measures have been proposed, ranging from (i) controlling the design stage to (ii) licensing and auditing requirements for scoring systems to (iii) ex post measures by public bodies.

In this vein, some authors propose for the US an oversight by regulators, such as the Federal Trade Commission (under its authority to combat unfair trade practices) with the possibility of accessing scoring systems, testing hypothetical examples by IT experts, issuing impact assessments evaluating the system's negative effects, and identifying risk mitigation measures.[191]

For the EU, some scholars suggest that the enforcement apparatus of the GDPR should be harnessed and used by national data protection authorities to make use of algorithmic audits and data protection impact assessments in uncovering the causes of bias and enforcing adequate metrics of algorithmic fairness.[192]

Although (European) Data protection law can surely help to mitigate risks of unfair and illegal discrimination, the GDPR is, on the other hand, no panacea. As *Zuiderveen Borgesius* points out, this is due to the following plausible reasons:[193] First, data protection authorities have limited financial and human resources to take effective action. In addition, many authorities may also lack the necessary expertise to detect and/or evaluate algorithmic discrimination. Second, the GDPR only covers personal data, not the ML models themselves.[194] Third, the regulation is vaguely formulated, which makes it difficult to apply its norms. Fourth, a conflict between data protection and anti-discrimination law arises when the use of sensitive personal data is necessary for avoiding discrimination in data-driven decision models.[195] And fifth, even if data protection authorities are granted extensive powers of control, the black box problem[196] still remains. In this respect, *Lipton* reminds us that the whole reason we turn to machine learning rather than "handcrafted decision rules" is that "for many problems, simple, easily understood decision processes are insufficient."[197]

For all these reasons, data protection law is not a cure-all against discrimination. Rather, further research is needed on the extent to which data protection law can contribute to the fight against algorithmic discrimination, whether there are still deficiencies to be addressed by other areas of law (e.g. consumer law, competition law, and – when ADM systems are used by public bodies – by administrative law and criminal law), or whether we need completely new rules.

---

[190] *Wachter/Mittelstadt/Floridi*, (2017) 7(2) International Data Privacy Law 76. Cf. also, in this book, *Sancho*, Chapter 4.

[191] *Citron/Pasquale*, (2014) 89 Washington Law Review 1. For a detailed overview of the various regulatory proposals, see *Mittelstadt/Allo/Taddeo/Wachter/Floridi*, (2016) July-September Big Data & Society 1, at p. 13.

[192] *Hacker* (n. 183). Cf. also *Mantelero*, Regulating Big Data, (2017) 33(5) Computer Law & Sec. Rev. 584; *Wachter*, Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR, (2018) 34(3) Computer Law & Sec. Rev. 436; *Wachter/Mittelstadt*, A right to reasonable inferences: Re-thinking data protection law in the age of Big Data and AI, (2018) Columbia Business Law Review, forthcoming, https://ssrn.com/abstract=3248829.

[193] *Zuiderveen Borgesius* (n. 186), pp. 24 et seq.

[194] Cf. *supra*, F.II.2.

[195] *Žliobaitė/Custers*, Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models, (2016) 24 Artif Intell law 183.

[196] Cf. *supra*, B.IV.

[197] *Lipton*, The Myth of Model Interpretability, KDnuggets, April 27, 2015, http://www.kdnuggets.com/2015/04/model-interpretability-neural-networks-deep-learning.html.

III. Market Manipulation: The Case of Algorithmic Collusion

Algorithms, in their increasingly widespread use, raise concerns over anti-competitive behavior, as they enable companies to achieve and sustain collusion without any formal agreement or human interaction.[198] This applies in particular to *dynamic pricing algorithms*. As the OECD points out in a recent report, pricing algorithms are "fundamentally affecting market conditions, resulting in high price transparency and high-frequency trading that allows companies to react fast and aggressively."[199] In concrete terms, such algorithms provide companies with the ability to evaluate a wide range of information relevant to pricing, in particular information about the competitors' pricing behavior, the current demand situation, price elasticity, and a number of other factors. On this basis, companies can adjust their own prices for thousands of products automatically and adapt them to the respective market situation in (milli)seconds.

According to *Stucke/Ezrachi,*[200] the following scenarios for algorithmic collusion can be distinguished:

First, pricing algorithms can be used to enforce a previously agreed upon pricing arrangement. This was the case, for example, with the so-called poster cartel, which was prosecuted by both the US and UK authorities.[201]

Secondly, competitors may use the same pricing algorithm, which may be programmed to prevent competition. Again, competition law provides sufficient means to address such behavior: If companies exchange their algorithms with rivals, it is a clear violation of competition law. In addition, collusive behavior can also occur when competitors purchase similar algorithms and data sets from the same third party. In this scenario, a so-called "hub and spoke" cartel may exist where co-ordination is, willingly or not, caused by competitors using the same "hub" for developing their pricing algorithms.[202]

A particular problem arises in the third constellation, in which competing companies use their own algorithms and datasets without evidence of an agreement between them. In this case, too, the use of pricing algorithms can lead to a restriction of competition. The high market transparency and the homogeneity of products in online trading facilitate parallel behavior. This situation is exacerbated if profit-maximizing algorithms are used. As pricing algorithms "observe" each other's price strategies and react directly to them, it is likely that a higher anti-competitive price will prevail. Since algorithms react immediately to any price change, companies have little incentive to gain an advantage through

---

[198] *Stucke/Ezrachi*, Artificial Intelligence and Collusion: When Computers Inhibit Competition, University of Tennessee, Legal Studies Research Paper Series #267, 2015, https://ssrn.com/abstract=2591874; *Ezrachi/Stucke*, Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy, 2016; *id.*, Two Artificial Neural Networks Meet in an Online Hub and Change the Future (of Competition, Market Dynamics and Society), 2017, https://ssrn.com/abstract=2949434; *Mehra*, Antitrust and the Robo-Seller: Competition in the Time of Algorithms, (2016) 100 Minnesota Law Review, 1323; *Oxera*, When Algorithms Set Prices: Winners and Losers, 2017, https://www.oxera.com/publications/when-algorithms-set-prices-winners-and-losers/; *Woodcock*, The Bargaining Robot, CPI Antitrust Chronicle (May 2017), https://ssrn.com/abstract=2972228.

[199] *OECD*, Algorithms and Collusion: Competition Policy in the Digital Age, 2017, www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm, p. 51.

[200] *Stucke/Ezrachi* (n. 198).

[201] US Department of Justice (DOJ) 2015, Former E-Commerce Executive Charged with Price Fixing in the Antitrust Division's First Online Marketplace Prosecution, Justice News of the US Department of Justice, Office of Public Affairs, www.justice.gov/opa/pr/former-e-commerce-executive-charged-price-fixing-antitrust-divisions-first-online-marketplace. For the UK see https://www.gov.uk/government/news/cma-issues-final-decision-in-online-cartel-case.

[202] *Ezrachi/Stucke*, Virtual Competition, 2016, pp. 46 et seq. For the EU, cf. also the *Eturas* case where a booking system was employed as a tool for coordinating the actions of the firms; ECJ, 21.1.2016, case C-74/14 (*Eturas*), ECLI:EU:C:2016:42.

price undercutting. Recent studies show that this scenario is indeed very likely:[203] Autonomous pricing algorithms may independently discover that they can make the highest possible profit if they avoid price wars. As a result, they may learn to collude even if they have not been specifically instructed to do so, and even if they do not communicate with one another. This is particularly problematic because in most countries (including the United States and EU Member States) such "tacit" collusion – not relying on explicit intent and communication – is currently treated as not illegal.

In addition, autonomous pricing algorithms give rise to new problems with respect to liability,[204] auditing, and monitoring[205] as well as enforcement.[206]

The same is true for other forms of market manipulation, e.g. for high frequency trading strategies such as quote stuffing (i.e. creating a lag in data availability in order to enhance latency arbitrage opportunities) and spoofing (i.e. placing large orders to create the impression of large demand or supply for a security, with the intention of driving the prevailing market price in a particular direction).[207] Here, too, arises the problem of attribution: as algorithmic systems interact at higher levels of automation and connectivity,[208] it becomes increasingly difficult to trace their behavior to a particular human actor and/or company.

[203] *Calvano/Calzolari/Denicolo/Pastorello*, Artificial Intelligence, Algorithmic Pricing and Collusion (December 20, 2018), https://ssrn.com/abstract=3304991. In contrast, cf. also *Schwalbe*, Algorithms, Machine Learning, and Collusion (June 1, 2018), https://ssrn.com/abstract=3232631 ("problem of algorithmic collusion rather belongs to the realm of legal sci-fi").

[204] Cf. *Mehra* (n. 198) at pp. 1366 et seq. See also *supra*, B.III. and E.

[205] Cf. *Ezrachi/Stucke*, Algorithmic Collusion: Problems and Counter-Measures, OECD Roundtable on Algorithms and Collusion, May 31, 2017, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DAF/COMP/WD%282017%2925&docLanguage=En, p. 25: "Due to their complex nature and evolving abilities when trained with additional data, auditing these networks may prove futile. The knowledge acquired by a Deep Learning network is diffused across its large number of neurons and their interconnections, analogous to how memory is encoded in the human brain."

[206] For further information see *Ezrachi/Stucke*, Virtual Competition, 2016, pp. 203 et seq.

[207] *Fisher/Clifford/Dinshaw/Werle*, Criminal forms of high frequency trading on the financial markets, (2015) 9(2) Law and Financial Markets 113.

[208] For the connectivity problem, see *supra*, B.I. On the problem of autonomy, see *supra*, B.III. and E.

H. (International) Initiatives to Regulate AI and Robotics

I. Overview

The previous overview shows that the use of AI systems and smart robotics raises a number of unresolved ethical and legal issues. Despite these findings, there is currently not a single country in the world with legislation that explicitly takes into account the problematic characteristics of autonomous systems[209] in general. Apart from a few exceptions,[210] there are also no special rules for AI systems and smart robotics in particular.

Admittedly, many countries and sometimes also international/intergovernmental organizations have rules, laws, and norms that are relevant for AI and robotics – ranging from constitutional principles (rule of law, democracy),[211] human rights,[212] and (international) humanitarian law;[213] to administrative and criminal law protecting inter alia fair procedures;[214] to special laws that could help to mitigate the described problems such as data protection law, cybersecurity law, product safety and product liability law, competition law, consumer law; and many other fields. These laws, however, were not made with AI and smart robotics in mind. Accordingly, it is difficult to gauge to what extent existing legislation sufficiently regulates the negative implications of AI.

Since the beginning of 2017, many governments in the world have begun to develop national strategies for the promotion, development, and use of AI systems. Still, as *Tim Dutton* – a Canadian Senior Policy Advisor who regularly updates a summary of the different AI policies – observes, no two strategies are

---

[209] Cf. thereto *supra*, B.

[210] Special regulation exists above all for self-driving vehicles, drones, and high frequency trading. In the US, most of the states have either enacted legislation or executive orders governing self-driving vehicles; cf. *National Conference of State Legislatures*, Autonomous Vehicles State Bill Tracking Database, http://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx. In 2017, the House of Representatives passed a bill for a "Self Drive Act" which was supposed to lay out a basic federal framework for autonomous vehicle regulation but, ultimately, failed to be considered on the Senate floor. In the EU, the Regulation on Civil Aviation 2018/1139 addresses issues of registration, certification, and general rules of conduct for operators of drones – however, without regulating civil liability directly; cf. *Bertolini*, Artificial Intelligence and civil law: liability rules for drones, Study, commissioned by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs at the request of the JURI Committee, PE 608.848, November 2018. In addition, the EU enacted provisions on High Frequency Trading, explained in this book by *Spindler*, Chapter 7. Moreover, in France, the Digital Republic Act (Loi no. 2016-1321 du 7 octobre 2016 pour une République numérique), provides that, in the case of state actors taking a decision "on the basis of algorithms", individuals have a right to be informed about the "principal characteristics" of the decision-making system. For more details see *Edwards/Veale*, Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?, (2018) May/June IEEE Security & Privacy 46.

[211] Cf. for example *Council of Europe*, Ethical Charter (n. 63).

[212] Cf. *Council of Europe*, Algorithms and Human Rights, Study on the Human rights dimensions of automated data processing techniques and possible regulatory implications, Council of Europe study, DGI(2017)12, prepared by the Committee of Experts on Internet Intermediaries (MSI-NET), 2018; *Berkman Klein Center*, Artificial Intelligence & Human Rights: Opportunities and Risks, September 25, 2018.

[213] *Margulies*, The Other Side of Autonomous Weapons: Using Artificial Intelligence to Enhance IHL Compliance (June 12, 2018), https://ssrn.com/abstract=3194713.

[214] On AI and administrative law cf. *Oswald/Grace*, Intelligence, Policing and the Use of Algorithmic Analysis: A Freedom of Information-Based Study, (2016) 1(1) Journal of Information Rights, Policy and Practice, https://journals.winchesteruniversitypress.org/index.php/jirpp/article/view/16; *Cobbe*, Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making, August 6, 2018, https://ssrn.com/abstract=3226913; *Coglianese/Lehr*, (2017) 105 Georgetown Law Journal 1147; https://ssrn.com/abstract=2928293.

alike.[215] Instead, national (and international) initiatives focus on a wide variety of aspects, such as research and development programs, skills and education, data and digital infrastructure, technical standardization, AI-enhanced public services, ethics and inclusion, and sometimes also legal standards. Whereas some countries have laid down specific and comprehensive AI strategies (e.g. China, the UK, France), some are integrating AI technologies within national technology or digital roadmaps (e.g. Denmark, Australia), while still others have focused on developing a national AI R&D strategy (US).[216]

In the US, most notably, the government already relied heavily under the Obama administration on the liberal notion of the free market.[217] In its report "Preparing for the Future of Artificial Intelligence", published in October 2016,[218] the White House Office of Science and Technology Policy (OSTP) explicitly refrains from a broad regulation of AI research and practice. Instead, the report highlights that the government should aim to fit AI into existing regulatory schemes, suggesting that many of the ethical issues related to AI can be addressed through increasing transparency and self-regulatory partnerships.[219] The Trump administration, too, sees its role not in regulating AI and robotics but in "facilitating AI R&D, promoting the trust of the American people in the development and deployment of AI-related technologies, training a workforce capable of using AI in their occupations, and protecting the American AI technology base from attempted acquisition by strategic competitors and adversarial nations" – thus maintaining US American leadership.[220]

By contrast, the AI strategy of the European Union, published in April 2018,[221] focuses not only on the potential impact of AI on competitiveness but also on its social and ethical implications.

The following lines will provide a brief overview of the AI strategy of the EU (II.) and the efforts of the most important international organizations in this field (III.). This is followed by a short outline of individual and collective efforts of companies and industries/branches for self-regulation (IV.). Single national AI strategies, on the other hand, are not discussed here as this would go beyond the scope of this chapter.

---

[215] *Dutton*, An Overview of National AI Strategies, June 28, 2018, https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd. Cf. also the overview by *Thomas*, Report on Artificial Intelligence: Part I – the existing regulatory landscape, May 14, 2018, https://www.howtoregulate.org/artificial_intelligence/.

[216] *Delponte* (n. 117) p. 22

[217] For a detailed discussion of the various AI strategies in the US, the EU, and the UK, see *Cath/Wachter/Mittelstadt/Taddeo/Floridi*, Artificial Intelligence and the "Good Society": the US, EU, and UK approach, (2018) 24(2) Sci Eng Ethics 505.

[218] *Executive Office of the President National Science and Technology Council Committee on Technology*, (2016), Preparing for the future of artificial intelligence (OSTP report), Washington, DC, USA, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf. The report followed five workshops and a public request for Information, cf. OSTP report, p. 12.

[219] *OSTP report* (n. 218), p. 17.

[220] *Donald Trump*, Executive Order on Maintaining American Leadership in Artificial Intelligence, issued on February 11, 2019, https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/. Cf. also *Shepardson*, Trump administration will allow AI to "freely develop" in U.S.: official, Technology News, May 10, 2018, https://www.reuters.com/article/us-usa-artificialintelligence/trump-administration-will-allow-ai-to-freely-develop-in-u-s-official-idUSKBN1IB30F.

[221] *European Commission*, Communication "Artificial Intelligence for Europe", COM(2018) 237 final.

II. European Union

1. The European Parliament's Resolution of February 2017

In the European Union, it was above all the European Parliament (EP) that first developed a strategy for an EU-wide regulation of AI and robotics. In February 2017, the EP passed a resolution "with recommendations to the Commission on Civil Law Rules on Robotics".[222] The resolution calls for the creation of a "European Agency for Robotics and AI" consisting of regulators and external technical and ethical experts who could provide the "technical, ethical and regulatory expertise needed to support the relevant public actors, at both Union and Member State level, in their effort to ensure a timely, ethical and well-informed response to the new opportunities and challenges",[223] and could monitor robotics-based applications, identify standards for best practice and, where appropriate, recommend regulatory measures, define new principles, and address potential consumer protection issues.[224] Moreover, the resolution recommends introducing an EU-wide registration system for specific categories of advanced robots.[225]

Apart from that, the EP proposes to develop a Charter on robotics consisting of a code of ethical conduct for researchers and designers to "act responsibly and with absolute consideration for the need to respect the dignity, privacy and safety of humans".[226] In addition, the EP asks the European Commission to clarify the liability of industry and autonomous robots when harm or damages occur and to adopt new rules on liability if necessary.[227]

2. The European Economic and Social Committee's Opinion on AI As of May 2017

Shortly after the European Parliament published its resolution, the European Economic and Social Committee (EESC) presented an opinion on AI at the end of May 2017,[228] in which the Committee provided not only an in-depth analysis of different types and subfields of AI, but also general recommendations, including a human-in-command approach for "responsible AI". In this regard, the opinion identifies eleven areas where AI poses societal and complex policy challenges, namely the following: ethics, safety, privacy, transparency and accountability, work, education and skills, (in-)equality and inclusiveness, law and regulation, governance and democracy, warfare, and super-intelligence.

3. The European Commission's AI Strategy and The Work of the High-Level Expert Group on AI

On 25 April 2018, two weeks after 25 European countries had signed the Declaration of Cooperation on AI with the goal to build on "the achievements and investments of Europe in AI" and agreed to shape a European approach on AI,[229] the European Commission published its Communication "Artificial

---

[222] *European Parliament*, Resolution (n. 21). The resolution does not include unembodied AI. Instead, AI is understood as an underlying component of "smart autonomous robots". Critically, *Cath* et al. (n. 217).
[223] *European Parliament*, Resolution (n. 21), No. 16.
[224] *European Parliament*, Resolution (n. 21), No. 17.
[225] *European Parliament*, Resolution (n. 21), No. 2.
[226] *European Parliament*, Resolution (n. 21), p. 19.
[227] *European Parliament*, Resolution (n. 21), Nos. 49 et seq. For details regarding the recommendations of the EP relating to liability, cf. E.III.2. and E.IV.
[228] *European Economic and Social Committee*, Opinion, Artificial intelligence – The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society (own-initiative opinion), Rapporteur: Catelijne Muller, INT/806.
[229] Declaration "Cooperation on Artificial Intelligence", Brussels, April 10, 2018, https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence.

Intelligence for Europe".[230] The document – complemented by another communication of 7 December 2018[231] – outlines three pillars as the core of the proposed strategy: (i) boosting the EU's technological and industrial capacity and AI uptake across the economy, (ii) preparing for socio-economic changes brought by AI, and (iii) ensuring an appropriate ethical and legal framework based on the Union's values and in line with the Charter of Fundamental Rights of the EU.

To support the implementation thereof, the Commission established the "High-Level Expert Group on Artificial Intelligence"[232] (AI HLEG) and mandated it with the drafting of two documents in particular: (i) AI Ethics Guidelines that build on the work of the European Group on Ethics in Science and New Technologies[233] and of the European Union Agency for Fundamental Rights,[234] and (ii) Policy and Investment Recommendations. At the same time, the European AI Alliance,[235] an open multi-stakeholder platform with over 2700 members, was set up to provide broader input for the work of the AI HLEG.

At the end of 2018, the AI HLEG presented its first draft, "Ethics Guidelines for Trustworthy AI".[236] After an open consultation which generated feedback from more than 500 contributors, the AI HLEG published the final version at the beginning of April 2019.[237] The guidelines are neither an official document from the European Commission nor legally binding. They are also not intended as a substitute for any form of policy-making or regulation, nor to deter from the creation thereof.[238]

One of the main goals of the guidelines is to ensure that the development and use of AI follows a human-centric approach, according to which AI is not seen as a means in itself but as a tool to enhance human welfare and freedom. To this end, the AI HLEG propagates "trustworthy AI" which is (i) lawful, complying with all applicable laws and regulations; (ii) ethical, ensuring adherence to ethical principles and values; and (iii) robust, both from a technical and social perspective. The document aims to offer guidance on achieving Trustworthy AI by setting out in Chapter I fundamental rights and ethical principles AI should comply with. From those fundamental rights and principles, Chapter II derives seven key requirements (human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental wellbeing; and accountability), which then lead in Chapter III to a concrete but non-exhaustive assessment list to apply the requirements of Chapter II, offering AI practitioners guidance.

---

[230] *European Commission*, Communication "Artificial Intelligence for Europe", COM(2018) 237 final.
[231] *European Commission*, Communication "Coordinated Plan on Artificial Intelligence", COM(2018) 795 final.
[232] https://ec.europa.eu/digital-single-market/en/high-level-group-artificial-intelligence.
[233] *European Group on Ethics in Science and New Technologies*, Statement on Artificial Intelligence, Robotics and "Autonomous" Systems, Brussels, March 9, 2018, https://ec.europa.eu/info/news/ethics-artificial-intelligence-statement-ege-released-2018-apr-24_en.
[234] The European Union Agency for Fundamental Rights (FRA), an independent EU body funded by the EU budget, started a new project on "Artificial Intelligence, Big Data and Fundamental Rights" in 2018 with the aim of helping create guidelines and recommendations in these fields. Cf. https://fra.europa.eu/en/about-fra/introducing-fra.
[235] https://ec.europa.eu/digital-single-market/en/european-ai-alliance.
[236] *The European Commission's High Level Expert Group on Artificial Intelligence*, Draft "Ethics Guidelines for Trustworthy AI", Working Document for stakeholders' consultation, Brussels, December 18, 2018, https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai.
[237] *AI HLEG*, Ethics Guidelines For Trustworthy AI (Ethics Guidelines), Brussels, April 8, 2019, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477. Moreover, the AI HLEG published the document "A Definition of AI", cf. *supra*, note 17.
[238] *AI HLEG*, Ethics Guidelines (n. 237), p. 3.

4. Next Steps

Starting in June 2019, the European Commission will launch a targeted piloting, focusing in particular on the assessment list which the AI HLEG has drawn up for each of the key requirements.[239] The feedback on the guidelines will be evaluated by the end of 2019. Building on this evaluation, the AI HLEG will review and update the guidelines at the beginning of 2020. In parallel, the AI HLEG is also working on policy and investment recommendations on how to strengthen Europe's competitiveness in AI.

The work of the AI HLEG is accompanied by evaluations of the current EU safety and liability framework. To this end, the Commission intends, with the help of other expert groups, (i) to issue a guidance document on the interpretation of the Product Liability Directive in light of technological developments by mid-2019; and (ii) to publish, also by mid-2019, a report on the broader implications for, potential gaps in, and orientations for the liability and safety frameworks for AI, Internet of Things, and robotics.[240]

III. International Organizations

Beyond the European Union, several international organizations have also taken the initiative to reflect on the future legal framework for AI and robotics, such as the Council of Europe (1.), the OECD (2.), and the United Nations (3.).

1. Council of Europe

The Council of Europe has already dealt with AI systems in the past, particularly with regard to Big Data analyses and their implications for data protection law. In addition to the Data Protection Convention 108,[241] the Council of Europe adopted several guidelines and recommendations which are important for AI systems, especially on profiling,[242] Big Data,[243] and the police sector.[244]

Most recently, the Convention's Consultative Committee published a report on "Artificial Intelligence and Data Protection: Challenges and Possible Remedies" by *Alessandro Mantelero*,[245] as well as guidelines on Artificial Intelligence.[246] Apart from this, the Council of Europe also published a study on "Algorithms and Human Rights" prepared by the Committee of Experts on Internet Intermediaries[247]

---

[239] *European Commission*, Communication "Building Trust in Human-Centric Artificial Intelligence", COM(2019) 168 final, p. 7.

[240] *European Commission*, Communication "Artificial Intelligence for Europe", COM(2018) 237 final, p. 16.

[241] Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, European Treaty Series – No. 108.

[242] *Council of Europe*, The protection of individuals with regard to automatic processing of personal data in the context of profiling, Recommendation CM/Rec(2010)13 and explanatory memorandum, https://rm.coe.int/16807096c3.

[243] *Council of Europe*, Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data, T-PD(2017)01, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a

[244] *Council of Europe*, Practical guide on the use of personal data in the police sector, T-PD(2018)01, http://rm.coe.int/t-pd-201-01-practical-guide-on-the-use-of-personal-data-in-the-police-/16807927d5.

[245] *Council of Europe* (n. 114).

[246] *Council of Europe*, Guidelines on Artificial Intelligence and Data Protection, T-PD(2019)01, https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8.

[247] *Council of Europe*, Algorithms and Human Rights (n. 212).

and another study on "Discrimination, artificial intelligence, and algorithmic decision making" written by *Zuiderveen Borgesius*.[248]

In addition, at the end of 2018, the Council of Europe's European Commission for the Efficiency of Justice adopted a "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment".[249] The Charter is the first European instrument to set out five principles that should apply to the automated processing of judicial decisions and data, based on AI techniques – namely the principle of respect for fundamental rights, the principle of non-discrimination, the principle of quality and security, the principle of transparency, and the principle "under user control" which should ensure that users are informed actors and in control of the choices made.

2. OECD

The Organisation for Economic Cooperation and Development (OECD) has been working on AI for several years.[250] In 2018, it created an expert group (AIGO) to provide guidance in scoping principles for AI in society. The expert groups' aim is to help governments, business, labor, and the public maximize the benefits of AI and minimize its risks. The expert group plans to develop the first intergovernmental policy guidelines for AI, with the goal of presenting a draft recommendation to the next annual OECD Ministerial Council Meeting in May 2019.[251]

Moreover, the OECD is planning to launch in 2019 a policy observatory on AI, i.e. "a participatory and interactive hub which would bring together the full resources of the organization in one place, build a database of national AI strategies and identify promising AI applications for economic and social impact".[252]

3. United Nations

The United Nations (UN) has also been discussing the use of AI systems for some time. Since 2014, under the aegis of the Convention on Certain Conventional Weapons (CCW), experts have been meeting annually to discuss questions related to lethal autonomous weapon systems (LAWS).[253]

Since 2017, the "AI for Good" series is the leading UN platform for dialogue on AI. The 2018 Summit generated AI-related strategies and supporting projects connecting AI innovators with public and/or private sector decision makers. During the 2018 Summit, more than 30 UN agencies met to discuss their roles in AI and solidify the UN-wide partnership. The results are published in a report which outlines the diverse activities taking place across the UN system.[254]

---

[248] *Zuiderveeen Borgesius* (n. 186).

[249] *Council of Europe*, Ethical Charter (n. 63).

[250] http://www.oecd.org/going-digital/ai/oecd-initiatives-on-ai.htm.

[251] http://www.oecd.org/going-digital/ai/oecd-moves-forward-on-developing-guidelines-for-artificial-intelligence.htm.

[252] http://www.oecd.org/going-digital/ai/oecd-moves-forward-on-developing-guidelines-for-artificial-intelligence.htm.

[253] Cf. especially Report of the 2017 UN Group of Governmental Experts on Lethal Autonomous Weapons Systems, November 20, 2017, https://www.unog.ch/80256EDD006B8954/(httpAssets)/B5B99A4D2F8BADF4C12581DF0048E7D0/$file/2017_CCW_GGE.1_2017_CRP.1_Advanced_+corrected.pdf. Moreover, see the European Parliament's resolution of 12 September 2018 on autonomous weapon systems, P8_TA-PROV(2018)0341.

[254] https://www.itu.int/pub/S-GEN-UNACT-2018-1.

IV. Industry Initiatives and Self-regulation at International Level

Over the last few years, several initiatives – propelled by the individual and collective efforts of researchers, practitioners, companies, and industries – have emerged with the task of developing ethical principles, best practices, and codes of conducts for the development and use of AI systems and robots.

The following initiatives and organizations are particularly noteworthy: AI Now Institute,[255] Association for Computing Machinery (ACM) with its Committee on Professional Ethics[256] and the Public Policy Council,[257] the Asilomar Principles of the Future of Life Institute,[258] the Foundation for Responsible Robotics,[259] Googles AI Principles,[260] The Institute of Electrical and Electronics Engineers (IEEE) Global Initiative on Ethics of Autonomous and Intelligent Systems,[261] OpenAI,[262] Partnership on AI,[263] Software and Information Industry Association (SIIS),[264] and The World Economic Forum's Center for the Fourth Industrial Revolution[265] amongst several others.

Of the initiatives mentioned here, the principles developed by the IEEE are likely to be the most comprehensive and influential. The IEEE is the world's largest technical professional body that plays an important role in setting technology standards. The current version of the treatise "Ethically Aligned Design"[266] is composed of more than one hundred recommendations for technologists, policy makers, and academics. They represent the collective input of several hundred participants from six continents. The goal of "Ethically Aligned Design" is "to advance a public discussion about how we can establish ethical and social implementations for intelligent and autonomous systems and technologies, aligning them to defined values and ethical principles that prioritize human well-being in a given cultural context."[267]

Finally, it should be noted that international standard setting organizations are also currently in the process of developing guidance for AI systems. To this end, the International Electrotechnical Commission of the International Organization for Standardization (ISO) created in 2018 a committee on AI which will provide guidance to other committees that are developing AI applications.[268]

---

[255] https://ainowinstitute.org/.
[256] https://ethics.acm.org/2018-code-draft-2/.
[257] https://acm.org/public-policy/usacm.
[258] https://futureoflife.org/ai-principles/.
[259] http://responsiblerobotics.org.
[260] https://www.blog.google/technology/ai/ai-principles/.
[261] https://ethicsinaction.ieee.org/.
[262] https://openai.com/.
[263] https://www.partnershiponai.org/. The Partnership on AI is an industry-led, non-profit consortium set up by Google, Apple, Facebook, Amazon, IBM, and Microsoft in September 2016 to develop ethical standards for researchers in AI in cooperation with academics and specialists in policy and ethics. The consortium has grown to over 50 partner organizations.
[264] *SIIS*, Ethical Principles for Artificial Intelligence and Data Analytics, 2017, http://www.siia.net/LinkClick.aspx?fileticket=b46tNqJuiJA%3d&tabid=577&portalid=0&mid=17113.
[265] https://www.weforum.org/center-for-the-fourth-industrial-revolution/areas-of-focus.
[266] *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2, http://standards.ieee.org/develop/indconn/ec/autonomous_ systems.html.
[267] https://standards.ieee.org/industry-connections/ec/autonomous-systems.html?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined.
[268] https://iecetech.org/Technical-Committees/2018-03/First-International-Standards-committee-for-entire-AI-ecosystem.

Similar efforts are currently being made by the three European standards institutions CEN, CENELEC, and ETSI.

I. Governance of Algorithms: Regulatory Options

I. Should AI Systems and Robotics be Regulated by Ethics or Law?

While governments, international organizations, companies, and industries around the world have begun developing ethical guidelines and standards and started discussing the future legal framework for AI and robotics, there is currently no consensus on what concrete measures should be taken going forward.

Today, many efforts focus on developing ethical principles. However laudable this work may be, it should be clear that soft law as such will not suffice. The work on ethical principles and guidelines can lay the groundwork for subsequent legislation, providing orientation on the possible content of legal rules. However, the main problem is that ethical guidelines and self-regulatory initiatives by industries are non-binding.[269] In addition, these principles are often too abstract to provide detailed guidance. As *Ben Wagner* has pointed out, "[M]uch of the debate about ethics seems increasingly focused on companies avoiding regulation. Unable or unwilling to properly provide regulatory solutions, ethics is seen as the 'easy' or 'soft' option which can help structure and give meaning to existing self-regulatory initiatives."[270] Indeed, ethical guidelines and self-regulation should not be used as an escape from (hard) regulation.

II. General Regulation versus Sector-specific Regulation

This raises the difficult question of which AI and robotics applications and which sectors require regulation. AI and robotic systems are used in many different sectors and for many different purposes, and often they do not threaten fundamental values. An AI-based spam filter does not carry the same risks as an AI system used by courts to predict the recidivism of offenders.

Even for AI systems that make decisions about humans, the problems arising from the use of algorithms can be quite different depending on the type of algorithm used, its purpose, the field of application, and the actors involved. Accordingly, a one-size-fits-all approach would be inappropriate. Rather, policy makers and scholars should determine the need for legislative action sector-specifically, taking into account the different risks and legal interests at stake.

III. Guiding Questions For Assessing the Need to Regulate

In order to gauge the need for new rules in a particular sector, we could consider, according to *Paul Nemitz,*[271] the following questions:

First, policymakers might ask which rules apply in a particular sector, whether these rules apply to AI and robotics, and whether they address the challenges in a sufficient and proportional manner. Hence, before making a new law, we should first determine the scope of the applicable rules, their underlying principles and goals, their ability to be applied in a specific context, and whether they are apt to tackle the problems posed by intelligent machines. In this context, policymakers should also take into account whether a particular action is legal under the existing law only because the action is performed by a machine and not by a human being. If this is the case, we should consider codifying the following principle: that an action carried out by AI is illegal if the same action carried out by a human would be illegal.

---

[269] *Saurwein/Just/Latzer*, Governance of algorithms: options and limitations, (2015) 17(6) Info 35.

[270] *Wagner*, Ethics as an Escape from Regulation: From Ethics-Washing to Ethics-Shopping?, in: Hildebrandt (ed.), Being Profiling. Cogitas ergo sum, 2008, pp. 108 et seq.

[271] *Nemitz*, Constitutional democracy and technology in the age of artificial intelligence, (2018) Phil. Trans. R. Soc. A 376.

A second aspect would be to evaluate whether regulatory principles found in specific bodies of law should be generalized for intelligent machines. For example, in most areas of sensitive human-machine interaction, and in particular in the law on pharmaceuticals, there is not only a far-reaching obligation to test products and undergo an authorization procedure before placing the product on the market, but also an obligation to monitor the effects of the product on humans. As *Nemitz* points out, "AI may be a candidate for such procedures and obligations, both on a general level, and with specific mutations, if developed for or applied in specific domains."[272]

A third way to assess the risks of intelligent systems and the corresponding need for regulation is to carry out an *algorithmic impact assessment*.[273] In this regard, inspiration can be drawn from Art. 35(1) GDPR which requires a data protection impact assessment when a practice is "likely to result in a high risk to the rights and freedoms of natural persons", especially when using new technologies. The introduction of such an impact assessment – combined with the obligation to monitor the risks of intelligent systems during its use – could strengthen the necessary dialogue between companies and policymakers and at the same time help to implement a general culture of responsibility in the tech industry.[274]

IV. Level of Regulation: Global, International, National, or Regional?

Given that AI and robotic systems are technologies with a global impact, some argue for worldwide regulation.[275] According to *Turchin/Denkenberger*,[276] such regulations could take the form of a UN agency similar to the International Atomic Energy Agency but with much tighter and swifter control mechanisms equivalent to a world government designed specifically for AI and robotics. The creation of such an agency is, however, unlikely in view of the fact that the UN is currently receiving less support from its Member States and international politics. Of course, this does not rule out the possibility that non-global solutions could reach the global level, especially if an external transfer mechanism is added such as an international agreement, or if a system based on local solutions becomes an influential global player.

For the European Union, the question also arises at which level regulation should take place. Since many areas of law have already been harmonized, current EU legislation should be re-evaluated to ensure that it is fit for intelligent machines. Any other approach would inevitably lead to a patchwork of national legislation, hampering the development and deployment of these systems. In this vein, the European Parliament recently called in a resolution for an "internal market for artificial intelligence" and called on the Commission "to evaluate whether it is necessary to update policy and regulatory frameworks in order to build a single European market for AI".[277]

---

[272] *Nemitz* (n. 271), p. 11.

[273] *Reisman* et al. discuss "algorithmic impact assessments" in the US; *Reisman/Schultz/Crawford/Whittaker*, Algorithmic impact assessments: A practical framework for public agency accountability, AI Now Institute, 2018, https://ainowinstitute.org/aiareport2018.pdf.

[274] The added value of such an algorithmic impact assessment compared to the procedure under Art. 35 GDPR could lie especially in the fact that important aspects beyond data protection could be analyzed.

[275] Cf. for example *Elon Musk*, quoted by *Morris*, Elon Musk: Artificial Intelligence Is the "Greatest Risk We Face as a Civilization", 2017, http://fortune.com/2017/07/15/elon-musk-artificial-intelligence-2/.

[276] *Turchin/Denkenberger* (n. 74).

[277] *European Parliament*, Resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics, P8_TA-PROV(2019)0081, no. 119.

V. Instruments for Modernizing the Current Legal Framework

Legislators have a wide range of instruments at their disposal for adjusting and updating the current regulatory and institutional framework. These instruments include the following:

- *Regulation of research and development* by banning certain algorithms or systems,[278] by denying research funds to systems with a high risk of misuse,[279] and/or by requiring that certain normative or ethical standards be taken into account already at the development stage (legality/ethics by design, in particular audibility by design),[280] following the "privacy by design" approach well known in data protection law[281]
- *Premarket Approval* systems, requiring that certain algorithms slated for use in certain applications must undergo a testing phase and obtain approval from an agency before deployment,[282] and/or introducing an obligatory *algorithmic impact assessment*,[283] following the model of the data protection impact assessment as foreseen in Art. 35(1) GDPR
- *Monitoring and oversight* by regulatory bodies in order to safeguard against undue risks and harm to the public, especially *auditing mechanisms* for algorithms consisting of testing, validation and/or verification of system performance and impact, carried out by internal or external auditors[284]
- *Ex post regulation* by private enforcement, especially by introducing "notice-and-take-down" procedures[285] and/or by updating liability/tort law[286]
- *Co-regulation*, i.e. regulatory cooperation between state authorities and the industry, e.g. (i) by schemes allowing companies to certify algorithms or products on the basis of voluntary algorithmic accountability standards which could be developed by standard setting organizations;[287] (ii) by seals of quality; or (iii) by using the regulatory policy of the New Approach, which has been applied for many years in the area of EU product safety law,[288] creating a presumption of conformity if products comply with harmonized standards
- *Accompanying measures* such as (i) creating a (EU) regulatory agency for AI and robotics;[289] (ii) introducing ethical review boards to assess the potential damages and benefits to society; (iii) developing a framework for explainable AI (XAI), covering both transparency (simulatability, decomposability, algorithmic transparency) and interpretability (textual descriptions,

[278] Cf. for example Art. 22(1) GDPR (prohibition of fully automated decisions).

[279] This option is being considered in particular by the UK House of Lords Select Committee on AI; cf. *Thomas* (n. 215).

[280] Cf. *Dignum* et al., Ethics by Design: necessity or curse?, in: Conitzer/Kambhampati/Koenig/Rossi/Schnabel (eds.), AIES 2018 – Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 2018, pp. 60 et seq.; *Leenes/Lucivero*, Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design, (2014) 6(2) Law, Innovation and Technology 194, https://ssrn.com/abstract=2546759.

[281] *Cavoukian*, Privacy by Design: Take the challenge, 2009.

[282] *Tutt*, An FDA For Algorithms, (2017) 69 Administrative Law Review 83, https://ssrn.com/abstract=2747994.

[283] For the US cf. *Reisman/Schultz/Crawford/Whittaker*, Algorithmic impact assessments: A practical framework for public agency accountability, AI Now Institute, 2018, https://ainowinstitute.org/aiareport2018.pdf. For the EU cf. *Martini*, in this book, Chapter 3.

[284] *Adler/Falk/Friedler* et al., Auditing Black-box Models for Indirect Influence, 2016, http://arxiv.org/abs/1602.07043; *Diakopoulos*, Algorithmic accountability: Journalistic investigation of computational power structures, (2015) 3(3) Digital Journalism 398; *Kitchin*, (2017) 20(1) Information, Communication & Society 14; *Sandvig/Hamilton/Karahalios/Langbort* (n. 55).

[285] For the Notice and Take-Down (N&TD) procedure in the US see Section 512(c) of the US Digital Millennium Copyright Act (DMCA). For the EU, see Art. 14 E-Commerce Directive 2000/31/EC.

[286] See *supra*, E.III.

[287] Cf. in this respect the certification procedures envisaged in Art. 42 GDPR.

[288] *Busch*, Towards a "New Approach" in European Consumer Law: Standardisation and Co-Regulation in the Digital Single Market, (2016) Journal of European Consumer and Market Law 197.

[289] *European Parliament*, Resolution (n. 21), No. 16. For the US, cf. *Calo*, The Case for a Federal Robotics Commission, September 1, 2014, https://ssrn.com/abstract=2529151; *Brundage/Bryson*, Smart Policies for Artificial Intelligence, August 29, 2016, https://arxiv.org/abs/1608.08196.

visualizations, local explanations, examples),[290] especially in order to provide for ex ante/ex post explanations about the systems functionality and ex post explanations about specific decisions; (iv) creating a right to know whether a person is interacting with a human being or a machine and whether she/he is subject to automated decision making;[291] and (v) a right to opt out or withdraw from automated decision making[292]

- *Improving cooperation* between the public and private sector and academia in order to reinforce knowledge sharing and promote education and training for designers on ethical implications, safety and fundamental rights, as well as for consumers on the use of robotics and AI.

Which of these tools is best suited and which of these instruments should be combined following a "multi-level legislation" approach cannot be answered in general terms. Both the choice of the regulatory instrument and the intensity of intervention ultimately depend on the type of the algorithmic system, its area of application (especially whether the system is used in the public or private sector) and – last but not least – on the degree of risk and the legal interests at stake.

VI. A Plea for an Innovation-Friendly Regulation

AI and robotics are fast-developing technologies. Adopting statutes or treaties may take years or even decades, whereas technology develops quickly, outpacing any attempt at regulating it. This "pacing problem"[293] is exacerbated by the well-known "Collingridge dilemma",[294] according to which at the early stages of a new technology, regulation is difficult due to lack of information, while by the time a technology's undesirable consequences are discovered, it is so much entrenched in our daily lives and economy that any control faces resistance from users, developers, and investors.

As AI and robotic systems permeate our lives already to a large extent, the need to address these regulatory challenges is even more urgent.

In order to deal with these problems, many scholars have suggested specific regulatory tools that could be considered in the creation of a future regulatory framework for AI and robotics:

- Phrasing statutes and guidelines in a *technology-neutral way* in order to ensure equal treatment[295] and sustainable rules[296]
- Using *multi-level legislation*, especially by combining statutory rules with guidelines that can be adopted, evaluated, and amended easily by regulatory bodies[297]
- Enhancing flexibility through "*temporary regulation*" by using "*experimental legislation*";[298]

---

[290] Cf. regarding these different (sub)categories *Lipton* (n. 197).

[291] Cf. *AI HLEG*, Ethics Guidelines (n. 237), p. 34.

[292] Cf. *AI HLEG*, Ethics Guidelines (n. 237), p. 34.

[293] *Marchant/Allenby/Herkert* (eds.), The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem, 2011; *Hagemann et al*., Soft Law for Hard Problems: The Governance of Emerging Technologies in an Uncertain Future, 2018, https://ssrn.com/abstract=3118539, p. 24

[294] *Collingridge*, The Social Control of Technology, 1980, pp. 11 et seq.

[295] *Reed*, Taking Sides on Technology Neutrality, (2007) 4(3) SCRIPTed 263, http://heinonline.org/HOL/P?h=hein.journals/scripted4&i=281.

[296] *Greenberg*, Rethinking Technology Neutrality, (2016) 100 Minnesota Law Review 1495.

[297] *Koops*, Should ICT regulation be technology-neutral?, in: Koops et al. (eds), Starting Points for ICT Regulation, 2006, https://ssrn.com/abstract=918746.

[298] *Fenwick/Kaal/Vermeulen*, Regulation Tomorrow: What Happens When Technology is Faster Than the Law? (2017) 6(3) American University Business Law Review 561, http://www.aublr.org/wp-content/uploads/2018/02/aublr_6n3_text_low.pdf; *Guihot/Matthew/Suzor*, Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence (July 28, 2017). Vanderbilt Journal of Entertainment & Technology Law, Forthcoming, https://ssrn.com/abstract=3017004, p. 50.

- Creating special zones for empirical testing and development in the form of a living lab,[299] or "*regulatory sandboxes*"[300] in which the regulator provides selected firms wishing to bring innovative products or services to market with an opportunity to roll out and test them within a designated domain for a specified period, subject to monitoring and oversight by the relevant regulator but without being forced to comply with the applicable set of rules and regulations
- Creating a "Governance Coordination Committee" to "provide oversight, cultivate public debate, and evaluate the ethical, legal, social, and economic ramifications of (…) important new technologies"[301]
- Implementing "*feedback processes*" in a dynamic regulatory framework that facilitate the enhancement of information for regulation in order to "enable rule makers to adapt to regulatory contingencies if and when they arise because a feedback effect provides relevant, timely, decentralized, and institution-specific information ex-ante"[302] and
- Applying a *data-driven approach* that enables dynamic regulation in order to identify what, when, and how to regulate.[303]

All these innovative regulatory techniques (and more) should be considered to deal with the manifold problems of AI and robotic systems. Since the risks of these systems are highly context-specific, there is no one-size-fits-all solution. Instead, there is a need for a multi-level legislation and a mix of different regulatory tools. Therefore, attention should shift to a mixed approach of abstract and concrete rules which combines different governance measures that mutually enable and complement each other.

J. Outlook

These existing uncertainties call for further risk and technology assessment to develop a better understanding of AI systems and robotics, as well as their social implications, with the aim of strengthening the foundations for evidence-based governance. Collaboration with computer science and engineering is necessary in order to assess the potential drawbacks and benefits, identify and explore possible developments, and evaluate whether ethical and legal standards can be integrated into autonomous systems (ethics/legality by design). Likewise, expertise from economics, political science, sociology, and philosophy is essential to evaluate more thoroughly how AI technologies affect our society. Since technical innovations know no boundaries, an international perspective is required. In this respect, the initiatives at the European and international levels are important and laudable.

Regulators should consider not only the existing laws and their underlying principles and goals, but also the regulatory bodies involved in the various sectors, different codes of conducts and international standards, ethical guidelines, and much more. This multiplicity of perspectives and approaches requires, moreover, an oversight and coordination of various principles, rules, codes, and interests.

---

[299] The model for such a living lab is the "Robot Tokku" created by the Japanese government in the early 2000s; cf. *Pagallo*, LegalAIze: Tackling the Normative Challenges of Artificial Intelligence and Robotics Through the Secondary Rules of Law, in: Corrales/Fenwick/Forgó (eds.), New Technology, Big Data and the Law, 2017, pp. 281 et seq., at pp. 293 et seq.

[300] Cf. *UK Financial Conduct Authority*, Regulatory Sandbox Lessons Learned Report, 2017, https://www.fca.org.uk/publications/research/regulatory-sandbox-lessons-learned-report; *Cummings*, Regulatory Sandboxes: A Practice for Innovation That Is Trending Worldwide, ETHNews, March 1, 2017, https://www.ethnews.com/regulatory-sandboxes-a-practice-for-innovation-that-is-trending-worldwide.

[301] *Marchant/Wallach*, Coordinating Technology Governance, (2015) XXXI(4) Issues in Science and Technology, https://issues.org/coordinating-technology-governance/.

[302] *Kaal/Vermeulen*, How to Regulate disruptive Innovation – From Facts to Data, July 11, 2016, https://ssrn.com/abstract=2808044, p. 25.

[303] *Kaal/Vermeulen* (n. 302); *Roe/Potts*, Detecting new industry emergence using government data: a new analytic approach to regional innovation policy, (2016) 18 Innovation 373.

In this spirit, policymakers should avoid premature, innovation-inhibiting regulation – but rather promote research and development projects that are committed to fundamental human values. Whether current development requires regulation, or whether such regulation would be too early for the time being, is indeed an open question. There is no one-size-fits-all solution. Instead, the need for new rules must be evaluated for each sector and for every application separately, considering the respective risks and legal interests involved.

We may think not only of "soft law" guidelines and ethical codes by industry bodies; updated sets of rules using traditional methods of regulation such as research and development oversight, product licensing, auditing mechanisms, co-regulation, and/or ex post public or private enforcement; but also of new, more fluid regulatory tools such as (data driven) experimental legislation or regulatory sandboxes.

What is necessary is a multi-level approach, combining different governance measures that mutually enable and complement each other, in order to find the right balance between keeping up with the pace of change and protecting from the harm posed by AI and robotic systems, while creating at the same time a regulatory environment that avoids over-regulation but allows for innovation and further development. Above all, much more research and debate is required to determine which rules, if any, are needed.