8

Fostering the Common Good

An Adaptive Approach Regulating High-Risk AI-Driven Products and Services

*Thorsten Schmidt and Silja Voeneky**

I. INTRODUCTION

The risks based on AI-driven systems, products, and services are human-made, and we as humans are responsible if a certain risk materialises and damage is caused. This is one of the main reasons why States and the international community as a whole should prioritise governing and responsibly regulating these technologies, at least if high-risks are plausibly linked to AI-based products or services.[1] As the development of new AI-driven systems, products, and services is based on the need of private actors to introduce new products and methods in order to survive as part of the current economic system,[2] the core and aim of the governance and regulative scheme should not hinder responsible innovation by private actors, but minimize risks as far as possible for the common good, and prevent violations of individual rights and values – especially of legally binding human rights. At least the protection of human rights that are part of customary international law is a core obligation for every State[3] and is not dependent on the respective constitutional framework or on the answer as to which specific international human rights treaty binds a certain State.[4]

* Thorsten Schmidt and Silja Voeneky are grateful for the support and enriching discussions at Freiburg Institute for Advanced Studies (FRIAS). Thorsten Schmidt wants to thank Ernst Eberlein, and Silja Voeneky all members of the interdisciplinary FRIAS Research Group *Responsible AI* for valuable insights. Besides, Voeneky's research has been financed as part of the interdisciplinary research project *AI Trust* by the Baden-Württemberg Stiftung (since 2020). Earlier versions of parts of Sections II-IV of this Chapter have been published in S Voeneky, 'Human Rights and Legitimate Governance of Existential and Global Catastrophic Risks' in S Voeneky and G Neuman (eds), *Human Rights, Democracy and Legitimacy in a World of Disorder* (2018) 139 *et seq.* and S Voeneky, 'Key Elements of Responsible Artificial Intelligence: Disruptive Technologies, Dynamic Law' (2020) 1 *OdW* 9 *et seq.*

[1] This approach is part of the concept of 'Responsible AI'. In the following, we concentrate on a regulative approach for high-risk AI-driven products; we nevertheless include – for a regulation *mutatis mutandis* – AI-based high-risk services.

[2] J Beckert and R Bronk, 'An Introduction to Uncertain Futures' in J Beckert and R Bronk (eds), *Uncertain Futures: Imaginaries, Narratives, and Calculation in the Economy* (2018), who link this to the capitalist system, only, which seems like a too narrow approach.

[3] Human rights treaties do *not* oblige non-state actors, such as companies; however, States are obliged to respect, protect, and fulfill human rights and the due diligence framework can be applied in the field of human rights protection; *cf.* M Monnheimer, *Due Diligence Obligations in International Human Rights Law* (2021) 13 *et seq.*, 49 *et seq.*, 204 *et seq.* With regard to a human-rights based duty of States to avoid existential and catastrophic risks that are based on research and technological development, *cf.* S Voeneky, 'Human Rights and Legitimate Governance of Existential and Global Catastrophic Risks' in S Voeneky and G Neuman (eds), *Human Rights, Democracy and Legitimacy in a World of Disorder* (2018) 139, 151 *et seq.* (hereafter Voeneky, 'Human Rights and Legitimate Governance').

[4] It is still disputed, however, whether there is an obligation for States to regulate extraterritorial corporate conduct, *cf.* M Monnheimer, *Due Diligence Obligations in International Human Rights Law* (2021) 307 *et seq.* For a positive answer Voeneky, 'Human Rights and Legitimate Governance' (n 3) 155 *et seq.*

In this chapter, we want to spell out core elements of a regulatory regime for high-risk AI-based products and such services that avoid the shortcomings of regimes relying primarily on preventive permit procedures (or similar preventive regulation) and that avoid, at the same time, the drawbacks of liability-centred approaches. In recent times both regulative approaches failed in different areas to be a solid basis for fostering justified values, such as the right to life and bodily integrity, and protecting common goods, such as the environment. This chapter will show that – similar to regulating risks that stem from the banking system – risks based on AI products and services can be diminished if the companies developing and selling the products or services have to pay a proportionate amount of money into a fund as a financial guarantee after developing the product or service but before market entry. We argue that it is reasonable for a society, a State, and also the international community to adopt rules that oblige companies to pay such financial guarantees to supplement preventive regulative approaches and liability norms. We will specify what amount of money has to be paid based on the *ex-ante* evaluation of risks linked to the high-risk AI product or AI-based service that can be seen as proportionate, in order to minimize risks, but fostering responsible innovation and the common good. Lastly, we will analyse what kind of accompanying regulation is necessary to implement the approach proposed by us. *Inter alia*, we suggest that a group of independent experts should serve as an expert commission to assess the risks of AI-based products and services and collect data on the effects of the AI-driven technology in real-world settings.

Even though the EU Commission has recently drafted a regulation on AI (hereafter: Draft EU AIA),[5] it is not the purpose of this chapter to analyze this proposal in detail. Rather, we intend to spell out a new approach that could be implemented in various regulatory systems in order to close regulatory gaps and overcome disadvantages of other approaches. We argue that our proposed version of an *'adaptive' regulation* is compatible with different legal systems and constitutional frameworks. Our proposal could further be used as a blueprint for an international treaty or international soft law[6] declaration that can be implemented by every State, especially States with companies that are main actors in developing AI-driven products and services.

The term AI is broadly defined for this chapter, covering the most recent AI systems based on complex statistical models of the world and the method of machine learning, especially self-learning systems. It also includes systems of classical AI, namely, AI systems based on software already programmed with basic physical concepts (preprogrammed reasoning),[7] as a symbolic-reasoning engine.[8] AI in its various forms is a multi-purpose tool or general purpose technology

[5] Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on AI (Artificial Intelligence Act) and amending certain Union Legislative Acts' COM(2021) 206 final.

[6] See Section II.

[7] For a broad definition see as well the Draft EU AIA; according to this AI system "means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations or decisions influencing the environments they interact with." Article 3(1) and Annex I Draft EU AIA reads: "(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods."

[8] *Cf.* recently M Bhatt, J Suchan, and S Vardarajan, 'Commonsense Visual Sensemaking for Autonomous Driving: On Generalised Neurosymbolic Online Abduction Integrating Vision and Semantics' (2021) 299 *Artificial Intelligence Journal* https://doi.org/10.1016/j.artint.2021.103522. Here we are concerned with the concept of 'object permanence', in other words, the idea that "discrete objects continue to exist over time, that they have spatial relationships with one another (such as in-front-of and behind)", the understanding that objects, such as cars, continue to exist even if they disappear behind an obstacle; see also 'Is a Self-Driving Car Smarter Than a Seven-Month-Old?' *The Economist* (2021) www.economist.com/science-and-technology/is-it-smarter-than-a-seven-month-old/21804141.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

and a rapidly evolving, innovative key element of many new and possibly disruptive technologies applied in many different areas.[9] A recent achievement, for instance, is the merger of biological research and AI, demonstrated by the use of an AI-driven (deep-learning) programme that a company can use to determine the 3D shapes of proteins.[10] Moreover, applications of AI products and AI-based services exist not only in the areas of speech recognition and robotics but also in the areas of medicine, finance, and (semi-)autonomous cars, ships, planes, or drones. AI-driven products and AI-driven services already currently shape areas as distinct as art or weapons development.

It is evident that potential risks accompany the use of AI-driven products and services and that the question of how to minimize these risks without impeding the benefits of such products and services poses great challenges for modern societies, States, and the international community. These risks can be caused by actors that are not linked to the company producing the AI system as these actors might misuse an AI-driven technology.[11] But damages can also originate from the unpredictability of adverse outcomes (so-called off-target effects[12]), even if the AI-driven system is used for its originally intended purpose. Damage might also arise because of a malfunction, false or unclear input data, flawed programming, etc.[13] Furthermore, in some areas, AI services or products will enhance or create new systemic risks. For example, in financial applications[14] based on deep learning,[15] AI serves as a cost-saving and highly efficient tool and is applied on an increasingly larger scale. The uncertainty of how the AI system reacts in an unforeseen and untested scenario, however, creates new risks, while the large-scale implementation of new algorithms or the improvement of existing ones additionally amplifies already existing risks. At the same time, algorithms have the potential to destabilize the whole financial system,[16] possibly leading to dramatic losses depending on the riskiness and the implementation of the relevant AI-driven system.

[9] S Russel and P Novig, *Artificial Intelligence: A Modern Approach* (3rd ed., 2016), 1. S Voeneky, 'Key Elements of Responsible Artificial Intelligence – Disruptive Technologies, Dynamic Law' (2020) 1 *OdW* 9, 10–11 with further references (hereafter Voeneky, 'Key Elements of Responsible Artificial Intelligence') https://ordnungderwissenschaft.de/wp-content/uploads/2020/03/2_2020_voeneky.pdf; I Rahwan and others, 'Machine Behaviour' (2019) *Nature* 568, 477–486 (2019) www.nature.com/articles/s41586-019-1138-y; for the various fields of application *cf.* also W Wendel, 'The Promise and Limitations of Artificial Intelligence in the Practice of Law' (2019) 72 *Oklahoma Law Review* 21, 21–24, https://digitalcommons.law.ou.edu/olr/vol72/iss1/3/.

[10] This might be a tool to solve the so-called protein folding problem, *cf.* E Callaway, '"It Will Change Everything": DeepMind's AI Makes Gigantic Leap in Solving Protein Structures' (2020) 588 *Nature* 203 www.nature.com/articles/d41586-020-03348-4.

[11] M Brundage and others, 'The Malicious Use of Artificial Intelligence' (*Malicious AI Report*, 2018) https://maliciousaireport.com/ 17.

[12] For this notion in the area of biotechnology, *cf.* XH Zhang and others, 'Off-Target Effects in CRISPR/Cas9-Mediated Genome Engineering' (2015) 4 *Molecular Therapy: Nucleic Acids* https://doi.org/10.1038/mtna.2015.37; WA Reh, *Enhancing Gene Targeting in Mammalian Cells by the Transient Down-Regulation of DNA Repair Pathways* (2010) 22.

[13] *Cf.* C Wendehorst in this volume, Chapter 12.

[14] Such as high-frequency trading, deep calibration, deep hedging and risk management. High-frequency trading means the automated trading of securities characterized by extremely high speeds and high turnover rates; deep calibration means the fitting of a model to observable data of derivatives (calibration) by deep neural networks and deep hedging means the derivation of hedging strategies by the use of deep neural networks. For details on the topic of AI and finance, *cf.* M Paul, Chapter 21, in this volume.

[15] To list a few examples of this rapidly growing field, *cf.* J Sirignano and R Cont, 'Universal Features of Price Formation in Financial Markets: Perspectives from Deep Learning' (2019) 19(9) *Quantitative Finance* 1449–1459; H Buehler and others, 'Deep Hedging' (2019) 19(8) *Quantitative Finance* 1271–1291; B Horvath, A Muguruza, and M Tomas, 'Deep Learning Volatility: A Deep Neural Network Perspective on Pricing and Calibration in (Rough) Volatility Models' (2021) 21(1) *Quantitative Finance* 11–27.

[16] J Danielsson, R Macrae, and A Uthemann, 'Artificial Intelligence and Systemic Risk' (*Systemic Risk Centre*, 24 October 2019) www.systemicrisk.ac.uk/publications/special-papers/artificial-intelligence-and-systemic-risk.

Even more, we should not ignore the risk posed by the development of so-called superhuman AI: Because recent machine learning tools like reinforcement learning can improve themselves without human interaction and rule-based programming,[17] it seems to be possible for an AI system – as argued by some scholars – to create an improved AI system which opens the door to produce some kind of artificial Superintelligence or superhuman AI (or 'the Singularity').[18] Superhuman AI might even pose a global catastrophic or existential risk to humanity.[19] Even if some call this a science-fiction scenario, other experts predict that AI of superhuman intelligence will happen by 2050.[20] It is argued, as well, that an intelligence explosion might lead to dynamically unstable systems and it becomes increasingly easy for smarter systems to make themselves smarter[21] that finally, there can be a point beyond which it is impossible for us to make reliable predictions.[22] In the context of uncertainty and 'uncertain futures',[23] it is possible that predictions fail and risks arise from these developments faster than expected or in an unexpected fashion.[24] From this, we deduce that superhuman AI can be seen as a low probability, high impact scenario.[25] Because of the high impact, States and the international community should not ignore the risks of superhuman AI when drafting rules concerning AI governance.

II. KEY NOTIONS AND CONCEPTS

Before spelling out in more detail lacunae and drawbacks of the current specific regulation governing AI-based products and services, there is a need to define key notions and concepts relevant for this chapter, especially the notions of regulation, governance, and risk.

When speaking about governance and regulation, it is important to differentiate between legally binding rules on the one hand at the national, European, and international level, and non-binding soft law on the other hand. Only the former are part of the law and regulation *strictu sensu*.

[17] *See* Y LeCun and others, 'Deep Learning' (2015) 521 *Nature* 436–444 www.nature.com/nature/journal/v521/n7553/full/nature14539.html.

[18] The term 'the Singularity' was coined in 1993 by the computer scientist Vernon Vinge; he argued that "[w]ithin thirty years, we will have the technological means to create superhuman intelligence," and he concluded: "I think it's fair to call this event a singularity ('the Singularity' (. . .))." *See* V Vinge, 'The Coming Technological Singularity: How to Survive in the Post-Human Era' in GA Landis (ed), *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace* (1993) 11, 12.

[19] See also in this volume J Tallinn and R Ngo, Chapter 2. *Cf.* as well S Hawking, 'Will Artificial Intelligence Outsmart Us?' in S Hawking (ed), *Brief Answers to the Big Questions* (2018), 181; S Russel and P Novig, *Artificial Intelligence: A Modern Approach* (3rd ed., 2016) 1036 *et seq.*; S Bringsjord and NS Govindarajulu, 'Artificial Intelligence' in EN Zalta (ed), *The Stanford Encyclopedia of Philosophy* (2020) https://plato.stanford.edu/entries/artificial-intelligence/ 9; A Eden and others, *Singularity Hypotheses: A Scientific and Philosophical Assessment* (2013); A Al-Imam, MA Motyka, and MZ Jędrzejko, 'Conflicting Opinions in Connection with Digital Superintelligence' (2020) 9(2) *IAES IJ-AI* 336–348; N Bostrom, *Superintelligence* (2014) esp. 75 (hereafter N Bostrom, *Superintelligence*); K Grace and others, 'When Will AI Exceed Human Performance? Evidence from AI Experts' (2018) 62 *Journal of Artificial Intelligence Research* 729–754 https://doi.org/10.1613/jair.1.11222.

[20] See *e.g.*, R Kurzweil, *The Singularity Is Near* (2005) 127; for more forecasts, see Bostrom, *Superintelligence* (n 14) 19–21.

[21] E Yudkowsky, 'Artificial Intelligence as a Positive and Negative Factor in Global Risk' in N Bostrom, MM Ćirković (eds), *Global Catastrophic Risks* (2011) 341.

[22] M Tegmark, 'Will There Be a Singularity within Our Lifetime?' in J Brockman (ed), *What Should We Be Worried About?* (2014) 30, 32.

[23] See for this J Beckert and R Bronk, 'An Introduction to Uncertain Futures' in J Beckert and R Bronk (eds), *Uncertain Futures: Imaginaries, Narratives, and Calculation in the Economy* (2018) 1–38, 2 who argue that 'actors in capitalist systems face an open and indeterminate future'.

[24] As argued in E Yudkowsky, 'There's No Fire Alarm for Artificial General Intelligence' (*Machine Intelligence Research Institute*, 13 October 2017) https://intelligence.org/2017/10/13/fire-alarm/.

[25] Voeneky, 'Human Rights and Legitimate Governance' (n 3) 150.

The term international soft law is understood in this chapter to include rules that cannot be attributed to a formal legal source of public international law and that are, hence, not directly legally binding. However, as rules of international soft law have been agreed upon by subjects of international law (i.e. States, International Organizations (IO)) that could, in principle, create international law[26] these rules possess a specific normative force and can be seen as relevant in guiding the future conduct of States, as they promised not to violate them.[27] Therefore, rules of international soft law are part of top down rulemaking, (i.e. regulation), and must not be confused with (bottom up) private rulemaking by corporations, including the many AI related codes of conduct, as for example, the Google AI Principles.[28]

In the following, regulation means only top down law making by States at the national, and European level or by States and IOs at the international level. It will not encompass rulemaking by private actors that is sometimes seen as an element of so-called self-regulation. However, in the following, the notion of governance will include rules that are part of top-down lawmaking (e.g. international treaties and soft law) and rules, codes, and guidelines by private actors.[29]

Another key notion for the adaptive governance framework we are proposing is the notion of risk. There are different meanings of 'risk' and in public international law, there is no commonly accepted definition of the notion, it is unclear how and whether a 'risk' is different from a 'threat', a 'danger', or a 'hazard'.[30] For the sake of this chapter, we will rely on the following broad definition, according to which a risk is an unwanted event that may or may not occur,[31] that is, an unwanted hypothetical future event. This definition includes situations of uncertainty, where no probabilities can be assigned for the occurrence of damage.[32] A global catastrophic risk

[26] For a similar definition, see D Thürer, 'Soft Law' in R Wolfrum (ed), *Max Planck Encyclopedia of Public International Law* (2012) volume 9 271 para 8.

[27] On the advantages and disadvantages of 'standards' compared to 'regulation' see J Tate and G Banda, 'Proportionate and Adaptive Governance of Innovative Technologies: The Role of Regulations, Guidelines, Standards' (*BSI*, 2016) www.bsigroup.com/localfiles/en-gb/bis/innovate%20uk%20and%20emerging%20technologies/summary%20report%20-%20adaptive%20governance%20-%20web.pdf 14 (hereafter Tate and Banda, 'Proportionate and Adaptive Governance').

[28] AI Google, 'Artificial Intelligence at Google: Our Principles' https://ai.google/principles/.

[29] It is beyond the scope of this chapter to discuss bottom up rules drafted by corporations or NGOs in the area of AI.

[30] See G Wilson, 'Minimizing Global Catastrophic and Existential Risks from Emerging Technologies through International Law' (2013) 31 *Va Envtl LJ* 307, 310. Sometimes there is no differentiation made between threat, hazard, and risk, *see* OECD Recommendation OECD/LEGAL/040 of 6 May 2014 of the Council on the Governance of Critical Risks www.oecd.org/gov/risk/Critical-Risks-Recommendation.pdf. For details see Voeneky, 'Human Rights and Legitimate Governance' (n 3) 140 *et seq*.

[31] See SO Hansson, 'Risk' in EN Zalta (ed), *Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/entries/risk/. In a quantitative sense, risk can be defined through risk measures (be it relying on probabilities or without doing so). Typical examples specify risk as to the probability of an unwanted event that may or may not occur (value-at-risk); or as the expectation of an unwanted event that may or may not occur (expected shortfall). The expectation of a loss materialized by the unwanted event is the product of its size in several scenarios with the probability of these scenarios and thus specifies an average loss given the unwanted event. Many variants of risk measures exist, see for example AJ McNeil, R Frey, and P Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools-Revised Edition* (2015). Adaptive schemes rely on conditional probabilities whose theory goes back to T Bayes, 'An Essay Towards Solving a Problem in the Doctrine of Chances' (1764) 53 *Phil Transactions* 370. In the area of international law, the International Law Commission (ILC) stated that the 'risk of causing significant transboundary harm' refers to the combined effect of the probability of occurrence of an accident and the magnitude of its injurious impact, *see* ILC, 'Draft Articles on Prevention of Transboundary Harm from Hazardous Activities' (2001) 2(2) *YB Int'l L Comm* 152.

[32] For a different, narrower notion of risk, excluding situations of uncertainty ('uncertainty versus risk'), see CR Sunstein, *Risk and Reason: Safety, Law and the Environment* (2002)129; CR Sunstein, *Worst-Case Scenarios* (2007)146–147; RA Posner, *Catastrophe* (2004) 171. A judge of the International Court of Justice (ICJ), however, included 'uncertain risks' into the notion of risks, see ICJ, *Pulp Mills on the River of Uruguay (Argentina v Uruguay)*, Sep Op of Judge Cançado Trindade [2010] ICJ Rep 135, 159, 162; for a similar approach (risk as 'unknown dangers') see J Peel, *Science and Risk Regulation in International Law* (2010) 1.

shall be defined as a hypothetical future event that has the potential to cause the death of a large number of human beings or/and to cause the destruction of a major part of the earth; and an existential risk can be defined as a hypothetical future event that has the potential to cause the extinction of human beings on earth.[33]

When linking AI-driven products and services to high-risks, we understand high-risks as those that have the potential to cause major damages for protected individual values and rights (as life and bodily integrity) or common goods (as the environment or the financial stability of a State).

The question of which AI systems, products, or services constitute such high-risk systems is discussed in great detail. The EU Commission has presented a proposal in 2021 as the core element of its Draft EU AIA regulating high-risk AI systems.[34] According to the Draft EU AIA, high-risk AI systems shall include, in particular, human-rights sensitive AI systems, such as AI systems intended to be used for the biometric identification and categorization of natural persons, AI systems intended to be used for the recruitment or selection of natural persons, AI systems intended to be used to evaluate the creditworthiness of natural persons, AI systems intended to be used by law enforcement authorities as polygraphs, and AI systems concerning the area of access to, and enjoyment of, essential private services, public services, and benefits as well as the area of administration of justice and democratic processes, thereby potentially affecting the rule of law in a State (Annex III Draft EU AIA). Nevertheless, it is open for debate whether high-risk AI products and services might include as well, because of the possibility to cause major damages, (semi-)autonomous cars, planes, drones, and ships, and certain AI-driven medical products (such as brain–computer-interfaces, mentioned below) or AI-driven financial trading systems.[35]

Additionally, autonomous weapons clearly fall under the notion of high-risk AI products. However, AI-driven autonomous weapon systems constitute a special case due to the highly controversial ethical implications and the international laws of war (*ius in bello*) governing their development and use.[36]

Another particular case of high-risk AI systems are AI systems that are developed in order to be part of or constitute superhuman AI – some even classify these AI systems as global catastrophic risks or existential risks.

[33] For slightly different definitions, see N Bostrom, 'Superintelligence' (n 12) 115 (stating that '[a]n existential risk is one that threatens to cause the extinction of Earth-originating intelligent life or to otherwise permanently and drastically destroy its potential for future desirable development'; and N Bostrom and MM Ćirković, 'Introduction' in N Bostrom and MM Ćirković (eds), *Global Catastrophic Risks* (2008) arguing that a *global catastrophic* risk is a hypothetical future event that has the potential 'to inflict serious damage to human well-being on a global scale'.

[34] *Cf.* n 5.

[35] For a definition of high-risk AI products by the European Parliament (EP), *cf.* EP Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics, and related technologies (2020/2021(INL)), para 14: 'Considers, in that regard, that artificial intelligence, robotics and related technologies should be considered high-risk when their development, deployment and use entail a significant risk of causing injury or harm to individuals or society, in breach of fundamental rights and safety rules as laid down in Union law; considers that, for the purposes of assessing whether AI technologies entail such a risk, the sector where they are developed, deployed or used, their specific use or purpose and the severity of the injury or harm that can be expected to occur should be taken into account; the first and second criteria, namely the sector and the specific use or purpose, should be considered cumulatively.' www.europarl.europa.eu/doceo/document/TA-9-2020-10-20_EN.html#sdocta9.

[36] Autonomous weapons are expressly outside the scope of the Draft EU AIA, *cf.* Article 2(3).

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

III. DRAWBACKS OF CURRENT REGULATORY APPROACHES OF HIGH-RISK AI PRODUCTS AND SERVICES

To answer the most pressing regulative and governance questions concerning AI-driven high-risk products and such services, this chapter introduces an approach for responsible governance that shall supplement existing rules and regulations in different States. The approach, spelled out below in more detail, is neither dependent on, nor linked to, a specific legal system or constitutional framework of a specific State. It can be introduced and implemented in different legal cultures and States, notwithstanding the legal basis or the predominantly applied regulatory approach. This seems particularly important as AI-driven high-risk products and such services are already being used and will be used to an even greater extent on different continents in the near future, and yet the existing regulatory approaches differ.

For the sake of this chapter, the following simplifying picture might illustrate relevant general differences: some States rely primarily on a preventive approach and lay down permit procedures or similar preventive procedures to regulate emerging products and technologies;[37] they even sometimes include the rather risk-averse precautionary principle, as it is the case according to EU law in the area of the EU policy of the environment.[38] The latter intends to oblige States to protect the environment (and arguably other common goods) even in cases of scientific uncertainty.[39] Other States, such as the United States, in many sectors, avoid strict permit procedures altogether or those with high approval thresholds or avoid a strict implementation, and rather rely on liability rules that give the affected party, usually the consumer, the possibility to sue a company and get compensation if a product or service has caused damage.

Both regulative approaches – spelling out a permit or similar preventive procedures, with regard to high-risk products or services in the field of emerging technologies, or liability regimes to compensate consumers and other actors after they have been damaged by using a high-risk product – even if they are combined have major deficits and have to be supplemented. On the one hand, preventive permit procedures are often difficult to implement and might be easy to circumvent, especially in an emerging technology field. This was illustrated in recent years in different fields, including emerging technologies, as by the aircraft MAX 737 incidents[40] or the

[37] The Draft AIA by the EU Commission spells out a preventive approach and does not include any relevant liability rules. However, the Commission has announced the proposal of EU rules to address liability issues related to new technologies, including AI systems in 2022, *cf.* C Wendehorst, Chapter 12, in this volume.

[38] See for the precautionary principle as part of EU law: Article 191(2) Treaty on the Functioning of the European Union, OJ 2016 C202/47 as well as Commission, 'Communication on the Precautionary Principle' COM(2000) 1 final. The precautionary principle (or: approach) is reflected in international law in Principle 15 of the Rio Declaration which holds that: 'In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, *lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures* to prevent environmental degradation.' (Emphasis added), United Nations, 'Rio Declaration on Environment and Development' (UN Conference on Environment and Development, 14 June 1992) UN Doc A/CONF 151/26/Rev 1 Vol I, 3; *cf.* also M Schröder, 'Precautionary Approach/Principle' in R Wolfrum (ed), *Max Planck Encyclopedia of Public International Law* (2012) volume 8, 400, paras 1–5. In philosophy, there has been an in-depth analysis and defense of the principle in recent times, *cf.* D Steel, *Philosophy and the Precautionary Principle: Science, Evidence, and Environmental Policy* (2014).

[39] It is also argued that this principle shall be applied in all cases of scientific uncertainty and not only in order to protect the environment, *cf.* C Phoenix and M Treder, 'Applying the Precautionary Principle to Nanotechnology' (*CRN*, January 2004) http://crnano.org/precautionary.htm; N Bostrom, 'Ethical Issues in Advanced Artificial Intelligence' (2003) https://nickbostrom.com/ethics/ai.html 2.

[40] As shown in T Sgobba, 'B-737 MAX and the Crash of the Regulatory System' (2019) 6(4) *Journal of Space Safety Engineering* 299; D Scharper, 'Congressional Inquiry Faults Boeing and FAA Failures for Deadly 737 Max Plane Crashes' *NPR News* (16 September 2020) www.npr.org/2020/09/16/913426448/congressional-inquiry-faults-boeing-and-

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

motorcar diesel gate[41] cases. If this is the case, damage caused by products after they entered the market cannot be avoided. On the other hand, liability regimes that allow those actors and individuals who suffered damage by a product or service to claim compensation, have the drawback that it is unclear how far they prevent companies from selling unsafe products or services.[42] Companies rather seem to be nudged to balance the (minor and unclear) risk to be sued by a consumer or another actor in the future with the chance to make (major) profits by using a risky technology or selling a risky product or service in the present.

How standard regulatory approaches fail was shown, *inter alia*, by the opiate crisis cases[43] in the United States.[44] Even worse, an accountability gap is broadened if companies can avoid or limit justified compensatory payments in the end via settlements or by declaring bankruptcy.[45]

faa-failures-for-deadly-737-max-plane-cr, key mistakes in the regulatory process were: 'excessive trust on quantitative performance requirements, inadequate risk-based design process, and lack of independent verification by experts.' It is argued that similar failures can happen in many other places, see for example P Johnston and H Rozi, 'The Boeing 737 MAX Saga: Lessons for Software Organizations' (2019) 21(3) *Software Quality Professional* 4.

[41] C Oliver and others, 'Volkswagen Emissions Scandal Exposes EU Regulatory Failures' *Financial Times* (30 September 2015) www.ft.com/content/03cdb23a-6758-11e5-a57f-21b88f7d973f; M Potter, 'EU Seeks More Powers over National Car Regulations after VW Scandal' *Reuters* (27 January 2017) www.reuters.com/article/us-volkswagen-emissions-eu-regulations-idUSKCN0V51IO.

[42] With regard to the disadvantages of the US tort system, MU Scherer, 'Regulating Artificial Intelligence' (2016) 29 *Harvard Journal of Law & Technology* 353, 388, and 391.

[43] The opiate crisis cases in the United States show in an alarming way that insufficient and low threshold regulation that allows to prescribe and sell a high-risk product without reasonable limits cannot be outweighed *ex post* by a liability regime, even if damaged actors claim compensation and sue companies that caused the damage, *cf.* District Court of Cleveland County, *State of Oklahoma, ex rel. Hunter v Purdue Pharma LP*, Case No CJ-2017-816 (2019).

[44] Another example are the actions of oil drilling companies, as the oil drill technology can be seen as a high-risk technology. As part of the the so-called 2010 Deepwater Horizon incident British Petroleum (BP) has caused an enormous marine oil spill. In 2014, US District Court for the Eastern District of Louisiana ruled that BP was guilty of gross negligence and willful misconduct under the US Clean Water Act (CWA). The Court found the company to have acted 'recklessly' (*cf.* US District Court for the Eastern District of Louisiana, *Oil Spill by the Oil Rig 'Deepwater Horizon' in the Gulf of Mexico on April 20, 2010*, Findings of Fact and Conclusion of Law, Phase One Trial, Case 2:19-md-02179-CJB-SS (4 September 2014) 121–122). In another case Royal Dutch Shell (RDS) was sued as its subsidiary in Nigeria had caused massive environmental destruction; the Court of Appeal in The Hague ordered in 2021 that RDS has to pay compensation to residents of the region and begin the purification of contaminated waters (*cf.* Gerechtshof Den Haag, *de Vereniging Milieudefensie v Royal Dutch Shell PLC and Shell Petroleum Development Company of Nigeria LTD/Shell Petroleum Development Company of Nigeria LTD v Friday Alfred Akpan*, 29 January 2021); see E Peltier and C Moses, 'A Victory for Farmers in a David-and-Goliath Environmental Case' *The New York Times* (29 January 2021) www.nytimes.com/2021/01/29/world/europe/shell-nigeria-oil-spills.html.

[45] This, as well, the opioid crisis cases in the United States have shown. *Cf.* J Hoffmann, 'Purdue Pharma Tentatively Settles Thousands of Opioid Cases' *New York Times* (11 September 2019) www.nytimes.com/2019/09/11/health/purdue-pharma-opioids-settlement.html: 'Purdue Pharma (...) would file for bankruptcy under a tentative settlement. Its signature opioid, OxyContin, would be sold by a new company, with the proceeds going to plaintiffs'. In September 2021, a federal bankruptcy judge gave conditional approval to a settlement devoting potentially $10 billion to fighting the opioid crisis but will shield the company's former owners, members of the Sackler family, from any future lawsuits over opioids, see J Hoffmann, 'Purdue Pharma Is Dissolved and Sacklers Pay $4.5 Billion to Settle Opioid Claims' *New York Times* (1 September 2021) www.nytimes.com/2021/09/01/health/purdue-sacklers-opioids-settlement.html. Several US states opposed the deal and planned to appeal against it, *cf.* 'What is the bankruptcy "loophole" used in the Purdue Pharma settlement?' *The Economist* (3 September 2021) www.economist.com/the-economist-explains/2021/09/03/what-is-the-bankruptcy-loophole-used-in-the-purdue-pharma-settlement. See also the Attorney General of Washington's statement of 1 September 2021: "This order lets the Sacklers off the hook by granting them permanent immunity from lawsuits in exchange for a fraction of the profits they made from the opioid epidemic — and sends a message that billionaires operate by a different set of rules than everybody else".

IV. SPECIFIC LACUNAE AND SHORTCOMINGS OF CURRENT AI REGULATION

If we take a closer look at the existing specific regulation and regulatory approaches to AI-driven products and (rarely) services, specific drawbacks become apparent at the national, supranational, and international level. It would be beyond the scope of this chapter to elaborate on this in detail,[46] but some loopholes and shortcomings of AI-specific rules and regulations shall be discussed below.[47]

1. *EU Regulation of AI-Driven Medical Devices*

A first example is the EU Regulation on Medical Devices (MDR),[48] which governs certain AI-driven apps in the health sector and other AI-driven medical devices such as in the area of neurotechnology.[49] The amended MDR was adopted in 2017 and entered into force in 2021.[50] It lays down a so-called scrutiny process[51] for high-risk products (certain class III devices) only, which is a consultation procedure prior to market. It regulates, *inter alia*, AI-driven medical device brain stimulation products, for example, brain–computer-interfaces (BCIs). They are governed by the MDR even if there is no intended medical purpose;[52] thus, the MDR also governs consumer neurotechnology devices.

However, it is a major drawback that AI-driven neurotechnology devices are regulated by the MDR, but this law does not lay down a permit procedure to ensure safety standards and only spells out the less strict scrutiny process. In this aspect, the regulation of AI systems intended for brain stimulation in the EU differs significantly from the regulations governing the development of drugs and vaccines in the EU which lay down rules with significantly higher safety thresholds, including clinical trials and human subjects research.[53] Considering the risks because of the use

[46] For this section see Voeneky, 'Key Elements of Responsible Artificial Intelligence' (n 9) 9 et seq.

[47] This does not include a discussion of AI and data protection regulations. However, the European General Data Protection Regulation (GDPR) aims to protect personal data of natural persons (Article 1(1) GDPR) and applies to the processing of this data even by wholly automated means (Article 2(1) GDPR). See 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, in force since 25 May 2018, OJ 2016 L119/1. The GDPR spells out as well a 'right to explanation' regarding automated decision processes; *cf.* T Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in T Wischmeyer and T Rademacher (eds), *Regulating Artificial Intelligence* (2019) 75 and 89: Article 13(2)(f) and 14(2)(g) as well as Article 22 GDPR contain an obligation to inform the consumer about the 'logic involved' as well as 'the significance and the envisaged consequences of such processing for the data subject' but not a comprehensive right to explanation.

[48] Regulation (EU) 2017/745 of the European Parliament and of the Council of 05 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, OJ 2017 L117/1. Besides, AI-based medical devices fall within the scope of high-risk AI systems according to Article 6(1) in conjunction with Annex II (11) Draft EU AIA that explicitly refers to Regulation 2017/745, if such AI systems are safety components of a product or themselves products and subject to third party conformity assessment, *cf.* this Section 3(b).

[49] According to Article 2 MDR 'medical device' '(. . .) means any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes: (. . .)'. For exemptions see, however, Article 1(6) MDR.

[50] The amended MRD came into force in May 2017, but medical devices are subject to a transition period of three years to meet the new requirements. This transition period was extended until 2021 due to the COVID-19 pandemic, *cf.* Regulation (EU) 2020/561 of the European Parliament and of the Council of 23 April 2020 amending Regulation (EU) 2017/745 on medical devices, as regards the dates of application of certain of its provisions.

[51] *Cf.* Articles 54, 55, and 106(3), Annex IX Section 5.1, and Annex X Section 6 MDR.

[52] Annex XVI: '(. . .) 6. Equipment intended for brain stimulation that apply electrical currents or magnetic or electromagnetic fields that penetrate the cranium to modify neuronal activity in the brain. (. . .)'.

[53] §§ 21 *et seq*. *Arzneimittelgesetz* (AMG, German Medicinal Products Act), BGBl 2005 I 3394; Article 3(1) Regulation (EC) 726/2004 of the European Parliament and of the Council of 31 March 2004 laying down Community procedures

of brain–computer-interfaces to humans and their health and integrity, it is unclear why the regulatory threshold is different from the development and use of drugs. This is even more true if neurotechnology is used as a 'pure' consumer technology by individuals and does not have a particular justification for medical reasons. Besides, there is no regulation of neurotechnology at the international level, and so far, no international treaty obliges the States to minimize or mitigate the risks linked to the use of AI-driven neurotechnology.[54]

2. *National Regulation of Semi-Autonomous Cars*

A second example of sector-specific (top down) regulation for AI-driven products with clear disadvantages that entered already in force are the rules governing semi-autonomous cars in Germany. The relevant German law, the *Straßenverkehrsgesetz*, hereafter Road Traffic Act, was amended in 2017[55] to include new automated AI-based driving systems.[56] From a procedural point of view it is striking that the law-making process was finalized before the federal ethics commission had published its report on this topic.[57] The relevant § 1a (1) Road Traffic Act states that the operation of a car employing a highly or fully automated (this means level 3, but not autonomous (not level 4 and 5))[58] driving function is permissible, provided that the function is used for its intended purpose:

> *Der Betrieb eines Kraftfahrzeugs mittels hoch- oder vollautomatisierter Fahrfunktion ist zulässig, wenn die Funktion bestimmungsgemäß verwendet wird.*[59]

It is striking that the meaning of the notions 'intended purpose' is not laid down by the Road Traffic Act itself or by an executive order but can be defined by the automotive company as a

for the authorization and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency, OJ 2004 L 136/1.

[54] The AI recommendation drafted by the OECD, *cf.* OECD Recommendation, OECD/LEGAL/0449 of 22 May 2019 of the Council on Artificial Intelligence https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 are also insufficient in this respect due to their non-binding soft law character, in more detail Voeneky, 'Key Elements of Responsible Artificial Intelligence', 17 *et seq.* and this Section at 3 a. However, at least some States such as Chile and France are attempting to regulate this area of AI: as part of the Chilean constitutional reform, the current Article 19 of the *Carta Fundamental* is to be supplemented by a second paragraph that protects mental and physical integrity against technical manipulation; *cf.* on the current status of the legislative process: *Cámara dediputadas y diputados*, Boletín No 13827-19 for an English translation of the planned amendment see www.camara.cl/verDoc.aspx?prmID=14151&prmTIPO=INICIATIVA, Anexo 1, p. 14. Furthermore, the implementation of specific 'neurorights' is planned, *cf.* project Boletín No 13828-19. The French bioethics law (*Loi n° 2021-1017 du 2 août 2021 relative à la bioéthique*), which came into force at the beginning of August 2021, allows the use of brain-imaging techniques only for medical and research purposes (Articles 18 and 19), *cf.* www.legifrance.gouv.fr/jorf/id/JORFTEXT000043884384/.

[55] *Straßenverkehrsgesetz* (StVG), *cf.* Article 1 Achtes Gesetz zur Änderung des Straßenverkehrsgesetzes (8. StVGÄndG), BGBl 2017 I 1648.

[56] §§ 1a, 1b and § 63 Road Traffic Act. For an overview of the most relevant international, European, and national rules governing autonomous or automated vehicles, *cf.* E Böning and H Canny , 'Easing the Brakes on Autonomous Driving' (FIP 1/2021) www.jura.uni-freiburg.de/de/institute/ioeffr2/downloads/online-papers/FIP_2021_01_BoeningCanny_AutonomousDriving_Druck.pdf (hereafter Böning and Canny, 'Easing the Brakes').

[57] Germany, Federal Ministry of Transport and Digital Infrastructure, Ethics Commission, 'Automated and Connected Driving' (*BMVI*, June 2017), www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.html.

[58] An act regulating fully autonomous cars has been in force since 2021 and has changed the Road Traffic Act, see especially the new §§ 1 d-1g Road Traffic Act. For the draft, *cf.* German Bundestag, 'Entwurf eines Gesetzes zur Änderung des Straßenverkehrsgesetzes und des Pflichtversicherungsgesetzes (*Gesetz zum autonomen Fahren*)' (Draft Act for Autonomous Driving) (9 March 2021), Drucksache 19/27439 https://dip21.bundestag.de/dip21/btd/19/274/1927439.pdf.

[59] § 1a (1) Road Traffic Act.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

private actor producing and selling the cars.[60] Therefore, the Road Traffic Act legitimizes and introduces insofar the private standard-setting by corporations. This provision thus contains an 'opening clause' for self-regulation by private actors but is, as such, too vague.[61] This is an example of a regulatory approach that does not provide sufficient standards in the area of an AI driven product that can be linked to high risks. Hence, it can be argued that the § 1a (1) Road Traffic Act violates the *Rechtsstaatsprinzip*, rule of law, as part of the German Basic Law,[62] which states that legal rules must be clear and understandable for those whom they govern.[63]

3. *General AI Rules and Principles: International Soft Law and the Draft EU AI Regulation*

The question arises whether the lacunae mentioned before at the national and European level in specific areas of AI regulation can be closed by rules of international law (a) and the future regulation at the European level, that is, the 2021 Draft AIA (b).

a. International Regulation? International Soft Law!

So far, there does not exist an international treaty regulating AI systems, products, or services. Nor is such a regulation being negotiated. The aims of the States, having their companies and national interests in mind, are still too divergent. This situation differs from the area of biotechnology, a comparable innovative and as well potentially disruptive technology. Biotechnology is regulated internationally by the the Cartagena Protocol, an international treaty, and this international biotech regulation is based on the rather risk averse precautionary principle.[64] Since more than 170 States are parties to the Cartagena Protocol,[65] one can speak of an almost universal regulation, even if the United States, as a major player, is not a State party and not bound by the Cartagena Protocol. However, even in clear high-risk areas of AI development, such as the development and use of autonomous weapons, an international treaty is still lacking. This contrasts with other areas of high-risk weapons development, such as those of biological weapons.[66]

Nevertheless, as a first step, at least international soft law rules have been agreed upon that spell out the first general principles governing AI systems at the international level. The Organization for Economic Co-operation and Development (OECD) has issued an AI Recommendation in 2019 (hereafter OECD AI Recommendation).[67] Over 50 States have

[60] Böning and Canny, 'Easing the Brakes' (n 56).

[61] This seems true even if the description of the intended purpose and the level of automation shall be 'unambiguous' according to the rationale of the law maker, *cf.* German Bundestag, 'Entwurf eines Gesetzes zur Änderung des Straßenverkehrsgesetzes' (Draft Act for Amending the Road Traffic Act) (2017), Drucksache 18/11300 20 https://dip21.bundestag.de/dip21/btd/18/113/1811300.pdf: *'Die Systembeschreibung des Fahrzeugs muss über die Art der Ausstattung mit automatisierter Fahrfunktion und über den Grad der Automatisierung unmissverständlich Auskunft geben, um den Fahrer über den Rahmen der bestimmungsgemäßen Verwendung zu informieren.'*

[62] Grundgesetz für die Bundesrepublik Deutschland (GG), BGBl 1949 I 1, last rev 29 September 2020, BGBl 2020 I 2048.

[63] B Grzeszick, 'Article 20' in T Maunz und G Dürig (eds), *Grundgesetz-Kommentar* (August 2020), para 99. This is not the case, however, with regard to level 4 and 5 autonomous cars, as the rules enshrined in the 2021 §§ 1 d-1 g Road Traffic Act are more detailed, even including some norms for a a solution of the so-called trolley problem, *cf.* § 1 e para. 2 (no 2).

[64] *Cf.* Section III.

[65] Cartagena Protocol on Biosafety to the Convention on Biological Diversity (adopted 29 January 2000, entered into force 11 September 2003) 2226 UTNS 208.

[66] Convention on the prohibition of the development, production, and stockpiling of bacteriological (biological) and toxin weapons and on their destruction (adopted 10 April 1972, entered into force 26 March 1975) 1015 UNTS 163.

[67] OECD AI Recommendation (n 54).

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

agreed to adhere to these principles, including States especially relevant for AI research and development, such as the United States, the UK, Japan, and South Korea. The OECD AI Recommendation states and executes five complementary value-based principles:[68] these are inclusive growth, sustainable development, and well-being (IV. 1.1); human-centred values and fairness (IV. 1.2.); transparency and explainability (IV. 1.3.); robustness, security, and safety (IV. 1.4.); and accountability (IV. 1.5.). In addition, AI actors – meaning those who play an active role in the AI system lifecycle, including organizations and individuals that deploy or operate AI[69] – should respect the rule for human rights and democratic values (IV. 1.2. lit. a). These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognized labor rights.

However, the wording of the OECD soft law principles is very soft ('should respect'). Even the OECD AI Recommendation on transparency and explainability (IV. 1.3.) has little substance. It states that

> [...] [AI Actors][70] should provide meaningful information, appropriate to the context, and consistent with the state of art: [...]
>
> to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

Assuming that discrimination and unjustified biases are one of the key problems of AI systems,[71] asking for a 'systematic risk management approach' to solve these problems,[72] seems insufficient as a standard of AI actors' due diligence.

Moreover, the OECD AI Recommendation does not mention any legal liability or legal responsibility. AI actors 'should be accountable'. This indicates that these actors should report and provide certain information about what they are doing to ensure 'the proper functioning of AI systems' and 'for the respect of the above principles' (IV. 1.5). This does not imply any legal obligation to achieve these standards or any legal liability if an actor fails to meet the threshold.

Finally, the OECD AI Recommendation does not stress the responsibility of governments to protect human rights in the area of AI. They include only five recommendations to policymakers of States ('adherents', section 2) that shall be implemented in national policies and international cooperation consistent with the above-mentioned principles. These include investing in AI research and development (V. 2.1), fostering a digital ecosystem for AI (V. 2.2), shaping and enabling policy environment for AI (V. 2.3), building human capacity and preparing for labour market transformation (V. 2.4), and international cooperation for trustworthy AI (V. 2.5). Hence, even if an actor aims to rely on the OECD AI Recommendation, it remains unclear what State obligations follow from human rights with regard to the governance of AI.

[68] An AI system is defined as 'a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.' *Cf.* OECD AI Recommendation (n 54).

[69] OECD AI Recommendation (n 54).

[70] AI actors here are 'those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI', see OECD AI Recommendation (n 54).

[71] See Data Ethics Commission, Opinion of the Data Ethics Commission (*BMJV*, 2019), 194 www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.pdf?__blob=publicationFile&v=3.

[72] 'AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.' *Cf.* IV. 1.4. c) OECD AI Recommendation (n 54).

Besides this, the problem of how to frame the low probability/high risk scenarious (or the low probability/catastrophic or existential risk challenges) linked to the possible development of superhuman AI is not even mentioned in the OECD AI Recommendation.[73]

b. Draft EU AI Regulation

As mentioned above, the draft regulation issued by the European Commission, the Draft EU AIA, proposes harmonized rules on AI systems and spells out the framework for general regulation of AI. It is laying down criteria with regard to requirements for the design and development of high-risk AI systems, not limited to specific sectors. For this, the regulation follows a risk-based regulatory approach – however not based on the precautionary principle – and, at its core, includes a classification of high-risk AI systems, on the one hand, and non-high-risk AI systems, on the other hand. For this, the notion of an AI system is defined in broad terms (Article 3(1) Draft EU AIA).[74] Also, the regulation governs all providers[75] 'placing on the market or putting into service AI systems in the EU' and all users of AI systems in the EU (Article 2, Article 3(2) Draft EU AIA). What kind of AI systems are high-risk AI systems, is laid down in general terms in Articles 6-7 and listed in Annex II and Annex III Draft EU AIA. The Annex III list, mentioned above,[76] can be amended and modified by the EU Commission in the future, which promises that the regulation might not be inflexible regulating the fast-moving field of AI systems as an emerging technology.[77]

The Draft EU AIA aims to limit the possible negative effects of the use of an AI system with regard to the protection of human rights, stressing core human rights as the protection of human dignity, autonomy, and bodily integrity. Therefore, certain 'AI practices' are prohibited according to Article 5 Draft EU AIA, especially if used by State authorities. This includes, but is not limited to, the use of certain AI systems that 'deploy[s] subliminal techniques beyond a person's consciousness' if this is likely to cause harm for a person. The same is true if AI practices cause harm to persons because they exploit the vulnerabilities of a specific group due to their age or disability, or the use of AI systems for law enforcement if this means to use a real-time remote biometric identification system. However, the latter prohibitions are not absolute as exemptions are enshrined in Article 5 Draft EU AIA.

Transparency obligations shall also protect human rights, as there is the need to make it transparent if an AI system is intended to interact with natural persons (Article 52 Draft EU AIA). The same is true with regard to the duty to report 'serious incidents or any malfunctioning (...) which constitutes a breach of obligations under Union law intended to protect fundamental rights' (Article 62 Draft EU AIA).

Apart from these prohibitions and duties, every high-risk AI system must comply with the specific requirements (Article 8 Draft EU AIA). This means that, *inter alia*, risk management systems must be established and maintained (Article 9 Draft EU AIA); training data sets must meet quality criteria (Article 10 Draft EU AIA). Besides, the criteria for the technical

[73] See note 36.

[74] See Section II.

[75] Providers are not limited to private actors but every natural or legal person, including public authorities, agencies, and other bodies, *cf.* Article 3(2).

[76] See Section II.

[77] The European Commission is entitled in Article 7 to add new high-risk systems to Annex III if those systems pose a risk to fundamental rights and safety that is comparable to those systems that are already contained in Annex III. However, this flexibility means that there is only a very loose thread of democratic legitimacy for the future amendments of Annex III. It is beyond the scope of this chapter to discuss this in more detail, but it is unclear whether this disadvantage is sufficiently justified because of the benefit to achieve more flexibility with regard to the regulation of AI systems as a fast-moving technology.

documentation of high-risk AI systems are spelled out in the Draft EU AIA (Article 11 and Annex IV); the operating high-risk AI systems shall be capable of the automatic recording of events and their operation has to be 'sufficiently transparent' (Article 12 and 13 Draft EU AIA). Finally, there must be human oversight (Article 14 Draft EU AIA); the latter could be interpreted as prohibiting the aim to develop and produce superhuman AI.

Another characteristic is that not only developing companies, providers of high-risk AI systems (Article 16 *et seq.* Draft EU AIA), importers and distributors (Articles 26 and 27 Draft EU AIA), but also users are governed by the Draft EU AIA and have obligations. Users encompass companies, as credit institutions, that are using high-risk AI systems (Articles 3(4), together with Articles 28 and 29 Draft EU AIA). Obligations are, for instance, that 'input data is relevant in view of the intended purpose of the high-risk AI system', and the duty to monitor the operation and keep the logs (Article 29 Draft EU AIA).

As the Draft EU AIA includes no relevant liability rules, it is a clear example of a preventive regulatory approach.[78] However, the Draft EU AIA does not establish a permit procedure but only a so-called conformity assessment procedure (Article 48 and Annex V Draft EU AIA), that is either based on internal control (Annex VI Draft EU AIA) or including the involvement of a notified body (Article 19 and 43, Annex VII Draft EU AIA). Notified bodies have to verify the conformity of high-risk AI systems (Article 33 Draft EU AIA). But it is up to the EU Member States to establish such a notifying authority (Article 30 Draft EU AIA) according to the requirements of the Draft EU AIA, and a notified body is allowed to subcontract specific tasks (Article 34 Draft EU AIA). As an oversight, the EU Commission can investigate cases 'where there are reasons to doubt' whether a notified body fulfills the requirements (Article 37 Draft EU AIA).

It has to be mentioned that derogations from the conformity assessment procedure are part of the regulation; derogations exist 'for exceptional reasons of public security or the protection of life and health of persons, environmental protection' and even (*sic!*) 'the protection of key industrial and infrastructure assets' (Article 47 Draft EU AIA).

In the end, many obligations rest on the providers, as for instance the documentation obligations (Article 50 Draft EU AIA), the post-market monitoring (Article 61 Draft EU AIA), or the registration of the system as part of the EU database (Articles 51 and 60 Draft EU AIA). However, if one evaluates how effective an implementation might be, it is striking that the regulation lays down only fines 'up to' a certain amount of money, as 10.000.000–30.000.000 EUR, if the Draft EU AIA is violated and it is up to the EU Member States to decide upon the severity of the penalties. Additionally, administrative fines that could be imposed on Union institutions, agencies, and bodies are much lower ('up to' 250.000 EUR – 500.000 EUR according to Article 72 Draft EU AIA).[79]

It is beyond the scope of this chapter to assess the Draft EU AIA in more detail.[80] Nevertheless, one has to stress that no permit procedure is part of the regulation of high-risk AI systems. This means that this regulation establishes lower thresholds with regard to high-risk AI systems compared, for instance, with the regulation of the development of drugs and vaccines in the EU. It seems doubtful whether the justification provided in the explanatory notes is convincing; it states that a combination with strong ex-post enforcement is an effective and

[78] For this differentiation, *cf.* Section III. For more details *cf.* C Wendehorst, Chapter 12, in this volume.
[79] For enforcement details *cf.* Articles 63 *et seq.*; for penalties *cf.* Article 71.
[80] For details *cf.* T Burri, Chapter 7, in this volume.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

reasonable solution, given the early phase of the regulatory intervention and the fact the AI sector is very innovative and expertise for auditing is only now being accumulated.[81]

In the end, without a regulative solution for liability issues, it seems doubtful whether the major risks of high-risk AI systems can be sufficiently mitigated on the basis of the Draft EU AIA. Therefore, another approach shall be proposed by us, one that is compatible with the Draft EU AIA but will complement it to fill in the loopholes.

4. Interim Conclusion

From what has been written above, one can conclude, firstly, that there are loopholes and drawbacks in the regulation of emerging technologies and especially AI systems, although there are rules in place in at least some areas of AI-driven products and services at the national, European, and international level. Secondly, there is no coherent, general, or universal international regulation of AI or AI-driven products and services.

Nevertheless, even outside the EU there is widespread agreement that there is the need to have proportional and robust regulation in place, at least for high-risk AI-driven products and such services. If we look at the multiple fields where AI-driven systems are currently used and could be used in the future and also look closely at the inherent benefits and risks linked to those systems and products it seems less surprising that prominent heads of companies selling AI-driven products have emphasized the urgent need to regulate AI systems, products, and services, as well.[82]

The vulnerability of automated trading systems on the financial market may serve as an example highlighting the huge impact of intelligent systems: In the Flash Crash 2010, a quickly completed order triggered automated selling, wiping out nearly $1,000 billion worth of US shares for a period of several minutes.[83]

Therefore, we agree with those who argue that high-risk AI products and such services are emerging and disruptive technologies that have to be regulated.[84] This is especially true with regard to high-risk AI services because these are often ignored. In our view, there is an urgent need for responsible, (i.e. robust) and proportional regulation of high-risk AI products and services today, because if we try to regulate these when major damages have already occurred, it will be too late.

[81] Critical on this as well C Wendehorst, Chapter 12, in this volume.

[82] This is true, for example, *Bill Gates*, *Sundar Pichai*, and *Elon Musk* have called for the regulation of AI. See S Pichai, 'Why Google Thinks We Need to Regulate AI' *Financial Times* (20 January 2020) www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04; E Mack, 'Bill Gates Says You Should Worry About Artificial Intelligence' (*Forbes*, 28 January 2015) www.forbes.com/sites/ericmack/2015/01/28/bill-gates-also-worries-artificial-intelligence-is-a-threat/; S Gibbs, 'Elon Musk: Regulate AI to Combat 'Existential Threat' before It's Too Late' *The Guardian* (17 July 2017) www.theguardian.com/technology/2017/jul/17/elon-musk-regulation-ai-combat-existential-threat-tesla-spacex-ceo: *Musk* stated in July 2017, at a meeting of the US National Governors Association, that 'AI is a fundamental risk to the existence of human civilization.'

[83] *Cf.* M Mackenzie and A van Duyn, '"Flash Crash" was Sparked by Single Order' *Financial Times* (1 October 2010) www.ft.com/content/8ee1a816-cd81-11df-9c82-00144feab49a. *Cf.* J Tallinn and T Ngo, Chapter 2, in this volume; M Paul, Chapter 21, in this volume.

[84] *Cf.* House of Lords Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able?* (Report of Session 2017–2019, 2018) HL Paper 100, 126 *et seq.*; MU Scherer 'Regulating Artificial Intelligence: Risks, Challenges, Competencies, and Strategies' (2016) 29(2) *Harvard Journal of Law & Technology* 353, 355; Perri 6, 'Ethics, Regulation and the New Artificial Intelligence, Part I: Accountability and Power' (2010) 4 *INFO, COMM & SOC'Y* 199, 203.

V. A NEW APPROACH: ADAPTIVE REGULATION OF AI-DRIVEN HIGH-RISK PRODUCTS AND SERVICES

1. *A New Approach*

We argue that a new approach to regulating AI-driven products is important to avoid the shortfalls of the rules at the national, supranational, and international level mentioned earlier. Our aim is to establish a regulatory approach that can supplement preventive procedures and, at the same time, close the gaps of liability-based approaches of different legal systems. This approach shall be applicable universally and could be laid down in national, supranational, or international laws. Our proposal aims for a proactive, adaptive regulatory scheme that is flexible, risk-sensitive, and has the incentive to assess and lower risks by those companies that develop and sell high-risk AI-driven products and such services. The proposal's core is that an operator or company must pay a proportionate amount of money (called regulatory capital in the following) as a financial security for future damages before a high-risk, AI-based product or such a service enters the market. To avoid over-regulation, we focus on AI-based products belonging to a class of high-risk products and services which, accordingly, have the potential to cause major damages for protected individual values, rights or interests, or common goods, such as life and bodily integrity, the environment, or the financial stability of a State. A regulatory framework for the potential development of superhuman AI will be discussed as well.

The special case of autonomous weapons, also a high-risk product, has to be mentioned as well: With regard to the specific problems of the development of (semi-)autonomous weapons, many authors and States state, based on convincing arguments, that a prohibition of these weapons is mandatory due to ethical and legal considerations.[85] This could mean that any kind of adaptive regulation suggested here should not be discussed as such regulation could be a safety net and justify the market entry of such weapons. We agree with the former, that a prohibition of such weapons is feasible, but disagree with the latter. Our argument for including (semi-)autonomous weapons in this discussion about responsible and adaptive regulation does not mean that we endorse the development, production, or selling of (semi-)autonomous weapons – quite to the contrary. Currently, however, it seems unlikely that there will be a consensus by the relevant States that develop, produce, or sell such weapons to sign an international treaty prohibiting or limiting these products in a meaningful way.[86] Therefore, this chapter's proposed regulatory approach

[85] As, for instance, the government of Austria, *cf.* Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, 'Proposal for a Mandate to Negotiate a Legally-Binding Instrument that Addresses the Legal, Humanitarian and Ethical Concerns Posed by Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS)' (Working Paper Submitted to the Convention on Conventional Weapons Group of Governmental Experts on Lethal Autonomous Weapons Systems by Austria, Brazil, and Chile, 8 August 2018) CCW/GGE.2/2018/WP.7 https://undocs.org/CCW/GGE.2/2018/WP.7; and *cf.* the decision of the *Österreicherischen Nationalrat*, Decision to Ban Killer Robots, 24 February 2021, www.parlament.gv.at/PAKT/VHG/XXVII/E/E_00136/index.shtml#.

[86] For the different State positions, see Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, 'Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (CCW)' (Report of the 2019 session of the GGE on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, 25 September 2019) CCW/GGW.1/2019/3 https://undocs.org/en/CCW/GGE.1/2019/3. On the discussion of these views *cf.* Voeneky, 'Key Elements of Responsible Artificial Intelligence' (n 9) 15–16. *Cf.* as well the resolution of the European Parliament, EP Resolution of 20 October 2020 with recommendations to the Commission on a framework for ethical aspects of artificial intelligence, robotics, and related technologies (2020/2012(INL)) www.europarl.europa.eu/doceo/document/TA-9-2020-10-20_DE.html#sdocta8.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

could, and should, at least close the responsibility gap that emerges if such weapons are developed and used. This seems to be urgently necessary as there are lacunae in the traditional rules of international humanitarian law,[87] and international criminal law,[88] and the international rules on State responsibility.[89] There is the danger that, because of these lacunae, States do not even have to pay compensation if, for instance, an autonomous weapon is attacking and killing civilians in clear violation of the rules of international law.

2. *Key Elements of Adaptive Regulation of AI High-Risk Products and Services*

We argue that adaptive regulation as a new regulatory scheme for AI-driven high-risk products and such services shall consist of the following core elements:

First, the riskiness of a specific AI-driven product or service should be evaluated by a commission of independent experts. The threshold regarding whether such an evaluation has to take place is dependent on whether the AI-based product or service falls into a high-risk category according to a *prima facie* classification of its riskiness that shall be laid down in legal rules.[90] Possible future scenarios together with available data on past experiences (using the evaluated or similar products or services) will form the basis for the experts' evaluation. If the evaluated product or service is newly developed, a certain number of test cases proposed by the expert commission should provide the data for evaluation.

Second, after the expert commission has evaluated whether a specific AI-driven product or service is high-risk as defined above and falls under the new regulatory scheme, and the questions are answered in the positive, the expert committee shall develop risk scenarios that specify possible losses and associated likelihoods for the scenarios to realize.

Third, relying, in addition to the riskiness of the product, on the financial situation of the developing or producing company,[91] the experts will determine the specific regulatory capital that has to be paid. They shall also spell out an evaluation system that will allow measurement and assessment of future cases for damages due to the implementation or operation of the AI-driven product or service.

Fourth, the set-up of a fund is necessary, into which the regulatory capital has to be paid. This capital shall be used to cover damages that are caused by the AI-driven high-risk product or service upon occurrence. After a reasonable time, for instance 5–10 years, the capital shall be paid back to the company if the product or service has caused no losses or damages.

Fifth, as mentioned above, after a high-risk product or service has entered the market, the company selling the product or service has to monitor the performance and effects of the product or service by collecting data. This should be understood as a compulsory monitoring phase in which monitoring schemes are implemented. The data will serve as an important source for future evaluation of the riskiness of the product by the expert commission.

[87] See Geneva Conventions (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 31, 85, 135, 287; Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3; Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II) (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 609.

[88] Rome Statute of the International Criminal Court (adopted 17 July 1998, entered into force 1 July 2002) 2187 UNTS 3.

[89] ILC, 'Materials on the Responsibility of States for Internationally Wrongful Acts' (United Nations, 2012) ST/LEG/SER.B/25.

[90] For a proposal by the EU Commission, cf. Section II.

[91] In contrast, the Draft EU AIA obliges 'providers' and 'users', see Section IV 3 b).

In particular, if the product or service is new and data is scarce, the evaluation system is of utmost importance because it serves as a database for future decisions on the amount of the regulatory capital and on the need for future monitoring of the product or service.

Sixth, another element of the proposed governance scheme is that the company should be asked to develop appropriate test mechanisms. A testing mechanism is a valid and transparent procedure ensuring the safety of the AI-driven product. For instance, a self-driving vehicle must pass a sufficient number of test cases to ensure that these vehicles behave in a safe way, meeting a reasonable benchmark.[92] Such a benchmark and test mechanism should be determined by the expert commission. Market entry should not be possible without a test mechanism in place. Given the data from the monitoring phase, the expert commission will be able to evaluate the product; but an appropriate test mechanism has additional advantages as the company itself can use it for the continuous evaluation of the product. It can support the re-evaluation explained in the next step. It will also help the regulator provide automized test mechanisms for the monitoring and evaluating of the technology, particularly in similar scenarios.

Seventh, the expert commission shall re-evaluate the AI-driven high-risk product or service on a regular basis, possibly every year. It can modify its decision on the proportionate amount of regulatory capital that is needed to match the risks by relying on new information and assessing the collected data. The established evaluation system mentioned above will provide reliable data for relevant decisions. (And, as mentioned earlier, after a reasonable time frame, the capital should be paid back to the company if the product or service has caused no losses or damages.)

3. *Advantages of Adaptive Regulation*

The following significant advantages follow from the adaptive approach[93] to regulation of AI high-risk prodocts and services: It avoids over-regulating the use of AI products and services especially in cases if the AI technology is new, and the associated risks are *ex ante* unclear. Current regulatory approaches that lay down preventive permit procedures can prevent a products' market entry (if the threshold is too high) or allow the market entry of an unsafe product (if the threshold is too low or is not implemented). With the adaptive regulation approach, however, it will be possible to ensure that a new AI product or AI-based service enters the market while sufficient regulatory capital covers possible future damages. The capital will be paid back to the company if the product or service proves to be a low-risk product or service after an evaluation period by using the data collected during this time according to the evaluation system.

a. Flexibility

The adaptive regulation approach allows reacting fast and in a flexible way to new technological developments in the field of AI. Since only the regulation's core elements are legally fixed *a priori*, and details shall be adapted on a case-by-case basis by an expert commission, the specific framing for an AI (*prima facie*) high-risk product can be changed depending on the information and data available. A periodical re-evaluation of the product or service ensures that new information can be taken into account, and the decision is based on the latest data.

[92] See, for example, T Menzel, G Bagschik, and M Maurer, 'Scenarios for Development, Test and Validation of Automated Vehicles' (2018) IEEE Intelligent Vehicles Symposium (IV).

[93] For the notion of adaptive governance *cf.* Tate and Banda, 'Proportionate and Adaptive Governance' (n 27) 4 *et seq.*, 20.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

b. Risk Sensitiveness

The approach is not only risk-sensitive with regard to the newly developed high-risk AI-based product or service; it also takes into account the different levels of risks accepted by different societies and legal cultures. It can be assumed that different States and societies are willing to accept different levels of risks linked to specific AI products and services, depending on the expected benefit. If, for instance, a society is particularly dependent on autonomous vehicles because of an ageing population and deficits in the public transport system, it might decide to accept higher risks linked to these vehicles to have the chance of an earlier market entry of the AI-based cars. According to these common aims, the threshold to enter the market laid down as part of a permit procedure could be lowered if, at the same time, the regulatory capital will be paid in the fonds and ensures that (at least) all damages will be compensated. The same is true, for instance, for AI-driven medical devices or other AI high-risk products that might be particularly important to people from one State and the common good of specific society due to certain circumstances.

c. Potential Universality and Possible Regionalization

Nevertheless, as AI systems are systems that could be used in every part of the world, the expert commission and its decision shall be based on international law. An international treaty, incorporating the adaptive regulation approach into international law, could outbalance lacunae or hurdles based on national admission procedures that might be ineffective or insufficient. The commission's recommendations or decisions, once made public, could be implemented directly in different national legal orders if the risk sensitiveness of the State is the same, and could serve as a supplement for the national admission process.

If, however, different types of risk attitudes towards an AI-driven high-risk product or such a service in different States exist, a cultural bias of risk averseness (or risk proneness) can be taken into account when implementing the proposal for regulation spelled out in this chapter at the national or regional levels. This allows the necessary flexibility of a State to avoid insufficient regulation (or overregulation) whilst protecting individual rights, such as bodily integrity or health, or promoting the common good, as the environment or the financial stability of a State or region. Such adjustments can be deemed necessary, especially in democratic societies, if risk perception of the population changes over time, and lawmakers and governments have to react to the changed attitudes. To that end, the German Constitutional Court (*Bundesverfassungsgericht*, BVerfG) has held that high-risk technologies (in the case at hand: nuclear energy) are particularly dependent on the acceptance of the population in the democratic society, because of the potentially severe damages that might be caused if they are used. The Constitutional Court stressed that because of a change in the public's perception of a high-risk technology, a reassessment of this technology by the national legislator was justified – even if no new facts were given.[94]

d. Monitoring of Risks

It can be expected that in most cases, a company producing a high-risk AI-driven product or service will be *a priori* convinced of the safety of its product or service and will argue that its AI-driven product or service can be used without relevant risks, while this opinion is possibly not

[94] BVerfGE 143, 246–396 (BVerfG *1 BvR 2821/11*) para 308. One of the questions in the proceedings was whether the lawmaker in Germany can justify the nuclear phase-out that was enacted after the reactor accident in Fukushima, Japan, took place. This was disputed as an 'irrational' change of German laws as the reactor accident in Fukushima did not, in itself, change the risk factors linked to nuclear reactors located in Germany.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

shared by all experts in the field. Therefore, the collection of data on the product's performance in real-world settings by the company evaluation systems is an important part of the adaptive regulation proposal introduced in this chapter. On the one hand, the data can help the company to show that its product or service is, as claimed, a low-risk product after a certain evaluation period and justify that the regulatory capital could be reduced or paid back; on the other hand, if the AI-driven product causes damages, the collected data will help improve the product and remedy future problems of using the technology. The data can also serve as an important source of information when similar products have to be evaluated and their risks have to be estimated. Hence, a monitoring phase is an important element of the proposal as reliable data are created on the product's or service's performance, which can be important at a later stage to prove that the technology is actually as riskless as claimed by the company at the beginning.

e. Democratic Legitimacy and Expert Commissions

The adaptive regulation approach spelled out in this chapter is not dependent on the constitution of a democratic, human rights-based State, but it is compatible with democracy and aims to protect core human and constitutional rights, such as life and health, as well as common goods, such as the environment. In order to have a sufficient basis that is legitimized, the rules implemented by the expert commission and the rules establishing the expert commission, should be based on an Act of parliament. Legally enshrined expert commissions or panels already exist in different contexts as part of the regulation of disruptive, high-risk products or technologies. They are a decisive element of permit procedures during the development of new drugs, as laid down for instance in the German Medicinal Products Act (*Arzneimittelgesetz*).[95] Another example of an interdisciplinary commission based on an act of parliament is the area of biotechnology regulation in Germany.[96]

As long as the commission's key requirements, such as the procedure for the appointment of its members, the number of members, the scientific background of members, and the procedure for the drafting of recommendations and decisions, are based on an act of parliament, a sufficient degree of democratic legitimacy is given.[97] In a democracy, this will avoid the pitfalls of elitism and an expert system, an expertocracy, that does not possess sufficient links to the legislature of a democratic State. A legal basis further complies with the requirements of human and constitutional rights-based constitutions, such as the German Basic Law, which demand that the main decisions relevant for constitutional rights have to be based on rules adopted by the legislative.[98]

[95] §§ 40(1), 42(1) AMG (n 53). For details *cf.* S Voeneky, *Recht, Moral und Ethik* (2010) 584–635, esp. at 594–606 (hereafter S Voeneky, *Recht, Moral und Ethik*).

[96] See the Central Committee on Biological Safety (ZKBS), an expert commission responsible for evaluating the risks concerning the development and use of genetically modified organisms (GMOs) www.zkbs-online.de/ZKBS/EN/Home/home_node.html. The commission is based on the the German Genetic Engineering Act (*Gentechnikgesetz* (GenTG)); BGBl 1993 I 2066 (§ 4 GenTG) and the decree, *Verordnung über die Zentrale Kommission für die Biologische Sicherheit* (ZKBS-Verordnung, ZKBSV) 30 October 1990 www.gesetze-im-internet.de/zkbsv/index.html.

[97] S Voeneky, *Recht, Moral und Ethik* (n 98).

[98] The so-called *Wesentlichkeitsprinzip*, that can be deduced from German Basic Law, is dependent on the constitutional framing and is not a necessary element of every liberal human rights-based democracy. In the United States, for instance, it is constitutional that the US president issues Executive Orders that are highly relevant for the exercise of constitutional rights of individuals, without the need to have a specific regulation based on an act of parliament. For the 'Wesentlichkeitsprinzip' according to the German Basic Law *cf.* S Voeneky, *Recht, Moral und Ethik* (2010) 214–218 with further references; B Grzeszick, 'Art. 20' in T Maunz und G Dürig (eds), *Grundgesetz-Kommentar* (August 2020) para 105.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

f. No Insurance Market Dependency

The adaptive regulation approach spelled out in this chapter avoids reliance on a commercial insurance scheme. An approach that refers to an insurance scheme that obliges companies to procure insurance for their AI-based high-risk products or services would depend on the availability of such insurances from companies. This could, however, fail for practical or structural reasons. Further, insurance might not be feasible for the development of new high-risk AI products and services if, and because, only a limited amount of data is available.[99] Besides, low probability-high-risk scenarios with unclear probability can hardly be covered adequately by insurances, as risk-sharing might be impossible or difficult to achieve by the insurer. Lastly, the reliance on insurance would mean that higher costs have to be covered by a company that is producing AI-based products, as the insurance company needs to be compensated for their insurance product and aims to avoid financial drawbacks by understating risks.

At the national level, there is an example that an attempt to regulate a disruptive technology, in this case biotechnology, based on the duty to get insurance failed as this duty was not implemented by either the regulator or the insurance industry.[100] Even at the international level, the duty to get insurance for operators can be seen as a major roadblock for ratifying and implementing an international treaty on the liability for environmental damage.[101]

4. *Challenges of an Adaptive Regulation Approach for AI-Driven High-Risk Products*

a. No Financial Means?

A first argument against the adaptive regulation approach could be that (different from financial institutions) the companies that develop and sell disruptive high-risk AI products or services do not have the capital to pay a certain amount as a guarantee for possible future damages caused by the products or service. This argument is, on the one hand, not convincing if we think about well-established big technology companies, like Facebook, Google, or Apple, etc., that develop AI products and services or outsource these developments to their subsidiaries.

On the other hand, start-ups, and new companies might develop AI-driven products and services which fall within the high-risk area. However, these companies often receive funding capital from private investors to achieve their goals even if they generate profit at a very late stage.[102] If an investor, often a venture capitalist, knows that the regulatory requirement is to pay a certain amount of capital to a fund that serves as security but that capital will be paid back to the company after a

[99] This is the problem existing with regard to the duty to get insurance for an operator that risks causing environmental emergencies in Antarctica as laid down in the Liability Annex to the Antarctic Treaty (Annex VI to the Protocol on Environmental Protection to the Antarctic Treaty: Liability Arising from Environmental Emergencies (adopted on 14 June 2005, not yet entered into force), *cf.* IGP&I Clubs, *Annex VI to the Protocol on Environmental Protection to the Antarctic Treaty: Financial Security* (2019), https://documents.ats.aq/ATCM42/ip/ATCM42_ip101_e.doc.

[100] Pursuant to § 36 GenTG (n 96) the German Federal Government should implement the duty to get insurance with the approval of the Federal Council (*Bundesrat*) by means of a decree. Such a secondary legislation, however, has never been adopted, *cf.* Deutscher Ethikrat, *Biosicherheit – Freiheit und Verantwortung in der Wissenschaft: Stellungnahme* (2014) 264 www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-biosicherheit.pdf.

[101] *Cf.* the so-called Liability Annex, an international treaty, not yet in force, that regulates the compensation of damages linked to environmental emergencies caused by an operator in the area of the Antarctic Treaty System, see note 58.

[102] For example, *Tesla* as a car manufacturer trying to develop (semi-)autonomous cars has only generated profit since 2020, *cf.* 'Tesla Has First Profitable Year but Competition Is Growing' (*The New York Times*, 27 January 2021) www.nytimes.com/2021/01/27/business/tesla-earnings.html.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

certain time if the product or service does not cause damages, this obligation would not impede or disincentivize the financing of the company compared to other requirements (for instance, as part of permit procedures). Quite to the contrary: To lay down a threshold of a certain amount of regulatory capital as a necessary condition before market-entry of an AI-based high-risk product (not for the stage of the research or development of the product) or AI-based service is an opportunity for the investor to take those risks into account that the company itself might downplay.

In the event that a State is convinced that a certain AI-driven product or service is fostering the common good of its society, and private investors are reluctant to finance the producing company because of major or unclear risks linked to the product or service, there is the possibility that the particular State may support the company with its financial means. Financial support has been given in different forms in other cases of the development of high-risk technology or products in the past and present.[103]

b. Ambiguity and Overregulation?

Another argument one could envisage against the adaptive regulatory approach introduced in this chapter is that it is unclear which AI-driven products or services have to be seen as high-risk products or high-risk services; and therefore there might be an inherent bias that leads to over-regulation as the category of high-risk products or services cannot be determined without grey areas, and can be determined neither precisely nor narrowly enough. However, what could be brought forward against this argument is that the category of high-risk AI products and services that the expert commission shall evaluate will be laid down in national, supranational, or international law after a process that includes the discourse with different relevant actors and stakholders, such as companies, developers, researchers, etc.[104] Criteria for a classification of prima facie high-risk AI products or services should be the possible damage that can occur if a certain risk linked to the product or service materializes. In order to avoid overregulation, one should limit the group of AI-driven high-risk products and services to the most evident; this might be depending on the risk proneness or risk awareness of a society as long as there is no international consensus.

c. Too Early to Regulate?

To regulate emerging technologies such as AI-based products and services is a challenge, and the argument is often brought forward that it is too early to regulate the technologies because the final product or service is unclear at a developmental stage. This is often linked to the argument that regulation of emerging technologies will mean inevitable overregulation of these technologies, as mentioned earlier. The answer to these arguments is that we as a society, every State, and the global community as a whole should avoid falling into the 'it is too early to regulate until it is too late' trap. Dynamic developments in a high-risk emerging technology sector, in particular, are characterized by the fact that sensible regulation rather might come too late, as legislative processes are, or can often be, lengthy. The advantage of the adaptive regulation proposed in this chapter is that, despite regulation, flexible standardization adapted to the specific case and the development of risk is possible.

[103] For instance, during the COVID-19 pandemic certain vaccine developing companies in Germany have been supported by the federal government and the EU; for example, the *Kreditanstalt für Wiederaufbau* (KfW) has acquired 'minority interest in CureVac AG on behalf of the Federal Government', *cf. KfW*, 6 August 2020. Also, the high-risk technology of nuclear power plants have been supported financially by different means in Germany since their establishment; *inter alia* the liability of the operating company in case of a maximum credible accident that has been part of the German law is capped and the German State is liable for compensation for the damages exceeding this cap, *cf.* §§ 25 *et seq.*, 31, 34, and 38 German Atomic Energy Act (*Atomgesetz* (AtG)), BGBl 1985 I 1565.

[104] *Cf.* above the proposals of the EU Parliament, note 35.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

d. No Independent Experts?

As mentioned earlier, the inclusion of expert commissions and other interdisciplinary bodies, such as independent ethics committees and Institutional Review Boards, has been established in various areas as an important element in the context of the regulation and assessment of disruptive, high-risk products or procedures. There are no reasons to assume why expert commissions should not be a decisive and important element in the case of AI regulation. Transparency obligations might ensure that experts closely linked to certain companies are not part of such a commission or are not part of a specific decision of such a commission. Moreover, a pluralistic and interdisciplinary composition of such a body is able to prevent biases as part of the regulative process.[105]

e. Unacceptable Joint Liability of Companies?

Further, it is not an argument against the fund scheme that companies that distribute AI-based products or services that later turn out to be low-risk are unduly held co-liable for companies that produce and distribute AI-based products or services that later turn out to be high-risk and cause damage. The aim of the fund's establishment is that claims for damages against a certain company X are initially compensated from the fund after a harmful case, namely from the sum that the harm-causing company X has deposited precisely for these cases concerning its risky AI products and services; should the amount of damage exceed this, further damages should initially be paid by company X itself. Thus, unlike with funds that contain a total capital that is depleted when damage payments are made in large amounts, it would be ensured that, in principle, the fund would continue to exist with the separate financial reserves of each company. If, to the contrary, the entire fund would be liable in the event of damage, the state where the company Y producing low-risk AI products is a national would have to provide a default liability to guarantee the repayment of the capital to the company Y. The state would be obliged to reimburse the paid-in regulatory capital to a company such as Y if, contrary to expert opinion, an AI product turns out to be low-risk and the regulatory capital has to be repaid to the company, but the fund does not have the financial means to do so due to other claims.

VI. DETERMINING THE REGULATORY CAPITAL

Central to the adaptive regulation proposed here is determining the level of regulatory capital. In this Section, we provide a formal setup, using probabilistic approaches.[106] In the first example, we consider a company that may invest in two competing emerging AI-driven products; one of the products is substantially riskier than the other. Even if we presume that the company is acting rationally (in the sense of a utility maximising[107] company),[108] there are good reasons to claim that risks exceeding the assets of the company will not be taken fully into account in the decision process of this company because, if the risks materialize, the bankruptcy of the company will be caused. Although it seems *prima facie* rational that diminishing risks exceeding the assets of the

[105] In the area of biotechnology *cf.* for instance in Germany the Central Committee on Biological Safety, ZKBS, note 96.

[106] *Cf.* VV Acharya and others, 'Measuring Systemic Risk' (2017) 30(1) *The Review of Financial Studies* 2–47 (hereafter Acharya and others, 'Measuring Systemic Risk').

[107] For this initial claim it is not necessary that utility is measured on a monetary scale. Later, when it comes to determining regulatory capital, we will, however, rely on measuring utility in terms of wealth.

[108] This means that future profits and losses are weighted with a utility function and then averaged by expectation. See for example DM Kreps, *A Course in Microeconomic Theory* (1990) or A Mas-Colell, MD Whinston, and JR Green, *Microeconomic Theory* (1995) volume 1.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

company should be the priority for the management of a company, as these risks threaten this actor's existence, the opposite behavior is incentivized. The high or even existential risks will be neglected by the company if there is no regulation in place obliging the company to take them into account: The company will seek high-risk investments because the higher return is not sufficiently downweighed by expected losses, which are capped at the level of the initial endowment.[109]

First Example: Two competing AI technologies or products

Consider a company with an initial endowment w_0. The company can decide to invest in two different AI-driven products or technologies offering (random) returns r and r' for the investment of 1 unit of currency. The first technology is the less risky one, while the second is riskier. We assume there are two scenarios: The first scenario (the best case, denoted by $+$) is if the risk does *de facto* not materialize. This scenario is associated with some probability p. In this scenario, the riskier strategy offers a higher return, i.e. $r(+) < r'(+)$.

In the second scenario (the worst case, denoted by $-$ and having probability $1-p$), the riskier technology will lead to larger losses, such that we assume $0 > r(-) > r'(-)$, both values being negative (yielding losses).

Summarizing, when the company invests the initial endowment into the strategy, the wealth at the end of the considered period (say at time 1) will be $w_1 = w_0 \cdot r$, on investing in the first technology, or $w_1' = w_0 \cdot r'$, when investing in the second, riskier technology, bankruptcy will occur when $w_1 < 0$, or $w_1' < 0$, respectively.

We assume that the company maximizes expected utility: Expected utility of the first strategy is given by the expectation of the utility of the wealth at time 1, $EU = E\left[u(w_1)1_{\{w_1>0\}}\right]$ (or $EU' = E\left[u(w_1')1_{\{w_1'>0\}}\right]$, respectively for the second strategy). Here u is a utility function[110] (we assume it is increasing), E denotes the expectation operator, and $1_{\{w_1>0\}}$ is the indicator function, being equal to one if $w_1 > 0$, (no bankruptcy) and zero otherwise (and similarly $1_{\{w_1'>0\}}$). The company chooses the strategy with the highest expected utility, namely, the first one if $EU > EU'$ and the second one if $EU' > EU$. If both are equal, one looks for additional criteria to find the optimal choice. This is typically a rational strategy.

Up to now, we have considered a standard case with two scenarios, a best case and a worst case. In the case of emerging and disruptive technologies, failure of high-risk AI systems and AI-driven products might lead to immense losses, such that in the worst-case scenario ($-$) bankruptcy occurs. This changes the picture dramatically:

> we obtain that $EU = p \cdot u(w_0 \cdot r(+))$ for the first technology, and for the second, riskier technology $EU' = p \cdot u(w_0 \cdot r'(+))$. Since the riskier technology's return in the best case scenario is higher, the company will prefer this technology. Most importantly, this does neither depend on the worst case's probability nor on the amount of the occurring losses. The company, by maximizing utility, will not consider losses beyond bankruptcy in its strategy.

Summarizing, the outcome of this analysis highlights the importance of regulation in providing incentives for the company to avoid overly risky strategies.

[109] See E Eberlein and DB Madan, 'Unbounded Liabilities, Capital Reserve Requirements and the Taxpayer Put Option' (2012) 12(5) *Quantitative Finance* 709–724 and references therein.

[110] A utility function associates to a various alternative a number (the utility). The higher the number (utility) is, the stronger the alternative is preferred. For example, 1 EUR has a different value to an individual who is a millionaire in comparison to a person who is poor. The utility function is able to capture such (and other) effects. See H Föllmer and A Schied, *Stochastic Finance: an Introduction in Discrete Time* (2011) Chapter 2 for further references.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

The first example highlights that a utility-maximising company will accept large risks surprisingly easily. In particular, the exact amount of losses does not influence the rational decision process, because losses are capped at the level of bankruptcy and the hypothetical losses are high enough to lead to bankruptcy regardless. It can be presumed that the company does not care about the particular amount of losses once bankruptcy occurs. This, in particular, encourages a high-risk strategy of companies since strategies with higher risk on average typically promise higher profits on average. However, the proposed adaptive regulation can promote the common good in aiming to avoid large losses. We will show below that the proposed regulation brings large losses back into the utility maximization procedure by penalizing high losses with high regulative costs, thus helping to avoid these.

Considering the problem of superhuman AI, a particular challenge arises: Once a company develops superhuman AI, the realized utility will be huge. It is argued that a superhuman AI cannot be controlled; thus, it is posing an existential threat not restricted to the company. Potential losses are clearly beyond any scale, yet any company will aim to develop such a superintelligent system as the benefits will be similarly beyond any scale.

The example highlights that a need for regulation will hopefully provide guidance for controlling the development of such AI systems when high-risk AI products lead to large losses and damages. However, with a low or even very low probability of this, large losses, once occurred, have to be compensated for by the public, since the company will be bankrupt and no longer able to cover them. Hence, regulation is needed to prevent a liability shortfall.

The following example will show that a reasonable regulation fosters an efficient maximization of overall wealth in comparison to a setting without regulation.

Second Example: A stylized framework for regulation

In this second example, regulatory capital is introduced. Adaptive regulation can maximize the overall wealth, minimize relevant risks, avoid large losses and foster the common good by requiring suitable capital charges.

Consider I companies: each company i has an initial wealth $\bar{w}_0^i$, where one part $\bar{w}_0^i - w_0^i$ is consumed initially, and the other part w_0^i is invested (as in the above example) resulting in the random wealth w_1^i at time 1. The company i pays a regulatory capital ρ^i and, therefore, aims at the following maximization:

$$max\Big[c \cdot \left(\bar{w}_0^i - w_0^i - \rho^i\right) + E\Big[u\Big(w_1^i 1_{\{w_1^i > 0\}}\Big)\Big]\Big]$$

The relevant rules should aim to maximize overall wealth: In the case of bankruptcy of a company, say i, the public and other actors have to cover losses. We assume that this is proportional to the occurred losses, $g \cdot w_1^i 1_{\{w_1^i < 0\}}$. The overall welfare function $P^1 + P^2$ consists of two parts: the first part is simply the sum of the utility of the companies,

$$P^1 = \sum_{i=1}^{I} c \cdot \left(\bar{w}_0^i - w_0^i - \rho^i\right) + E\Big[u\Big(w_1^i 1_{\{w_1^i > 0\}}\Big)\Big].$$

The second part,

$$P^2 = \sum_{i=1}^{I} E\Big[g \cdot w_1^i 1_{\{w_1^i < 0\}}\Big],$$

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

is the expected costs in case of bankruptcies of the companies. As scholars argue,[111] one obtains the efficient outcome, maximizing overall wealth or the common good, respectively, by choosing regulatory capital as

$$\rho^i = \frac{g}{c} \cdot P\left(w_1^i < 0\right) \cdot ES^i; \tag{1}$$

here the expected shortfall is given by $ES^i = -E\left[w_1^i 1_{\{w_1^i<0\}}\right]$. Hence, by imposing this regulatory capital, the companies will take losses beyond bankruptcy into account, which will help to achieve maximal overall wealth.

As spelled out in the literature, one could incorporate systemic effects in addition, which we do not consider here for simplicity.[112]

Here the adaptive regulatory approach relies on expectations and, therefore, assumes that probabilities can be assessed, even if they have to be estimated[113] or suggested by a team of experts. In the case of high uncertainty, this might no longer be possible, and one can rely on non-linear expectations (i.e. utilize *Frank Knight's* concept of uncertainty or in the related context of 'uncertain futures'). As already mentioned, the projection of unknown future risks can be formalized by relying on extreme value theory.[114] Therefore, it is central that adapted methods are used to incorporate incoming information resulting from the above mentioned monitoring process or other sources. The relevant mathematical tools for this exist.[115]

VII. DISSENT AND EXPERT COMMISSION

With regard to the expert commission, one has to expect that a variety of opinions arise. One possibility is that the worst-case opinion is considered, that is, taking the most risk-averse view. An excellent alternative to taking best-/worst-case scenarios or similar estimates is to rely on the underlying estimates' credibility. This approach is based on the so-called credibility theory, which combines estimates, internal estimates, and several expert opinions in the actuarial context.[116] We show how and why this is relevant for the proposed regulation.

[111] Acharya and others, 'Measuring Systemic Risk' (n 109).
[112] Acharya and others, ibid.
[113] M Pitera and T Schmidt, 'Unbiased Estimation of Risk' (2018) 91 *Journal of Banking & Finance* 133–145.
[114] See, for example L De Haan and A Ferreira, *Extreme Value Theory: An Introduction* (2007).
[115] See, for example AH Jazwinski, *Stochastic Processes and Filtering Theory* (1970), R Frey and T Schmidt, 'Filtering and Incomplete Information' in T Bielecki and D Brigo (eds), *Credit Risk Frontiers* (2011); T Fadina, A Neufeld, and T Schmidt, 'Affine Processes under Parameter Uncertainty' (2019), 4.1 *Probability, Uncertainty and Quantitative Risk*, 1–35.
[116] Credibility theory refers to a *Bayesian* approach to weight the history of expert opinions, see the recent survey by R Norberg (2015) 'Credibility Theory' in N Balakrishnan and others (eds) *Wiley StatsRef: Statistics Reference Online* or the highly influential work by H Bühlmann, 'Experience Rating and Credibility Theory' (1967) 4(3) *ASTIN Bulletin* 199.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press

Third Example: Regulation relying on credibility theory

For simplicity, i will be fixed, and we consider only two experts, one suggesting the probability P_1 and the other one P_2. The associated values of the regulatory capital computed using equation (1) are denoted by ρ_1 and ρ_2, respectively.

The idea is to mix ρ_1 and ρ_2 for the estimation of the regulatory capital as follows:

$$\rho^{credible}(\theta) = \theta \cdot \rho_1 + (1-\theta) \cdot \rho_2$$

where θ will be chosen optimal in an appropriate sense. If we suppose that there is already experience on estimates of the two experts, we can obtain variances v_1 and v_2 estimated from their estimation history. The estimator having minimal variance is obtained by choosing

$$\theta_{opt} = \frac{v_2}{v_1 + v_2}.$$

When expert opinions differ, credibility theory can be used to provide a valid procedure for combining the proposed models. Systematic preference is given to experts who have provided better estimates in the past. Another alternative is to select the estimate with the highest (or lowest) capital; however, this would be easier to manipulate. More robust variants of this method based on quartiles, for example, also exist.

VIII. SUMMARY

This chapter spells out an adaptive regulatory model for high-risk AI products and services that requires regulatory capital to be deposited into a fund based on expert opinion. The model allows compensating potentially occurring damage, while at the same time motivating companies to avoid major risks. Therefore, it contributes to the protection of individual rights of persons, such as life and health, and to the promotion of the common good, such as the protection of the environment. Because regulatory capital is reimbursed to a company if an AI high-risk product or service is safe and risks do not materialize for years, we argue that this type of AI regulation will not create unnecessarily high barriers to the development, market entry, and use of new and important high-risk AI-based products and services. Besides, the model of adaptive regulation proposed in this chapter can be part of the law at the national, European, and international level.

https://doi.org/10.1017/9781009207898.011 Published online by Cambridge University Press