The law and economics of AI liability

Buiten , Miriam C.; De Streel, Alexandre; Peitz, Martin

Published in:
Computer Law and Security Report

DOI:
10.1016/j.clsr.2023.105794

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

Comment

The law and economics of AI liability ☆

Miriam Buiten [a,b,*], *Alexandre de Streel* [b,c], *Martin Peitz* [b,d]

[a] *University of St. Gallen, Law School, Switzerland*
[b] *CERRE, Brussels, Belgium*
[c] *University of Namur and NADI, Belgium*
[d] *University of Mannheim, Department of Economics and MaCCI, Germany*

ARTICLE INFO

JEL classification:
K13
O33

Keywords:
Artificial intelligence
Liability
EU law

ABSTRACT

The employment of AI systems presents challenges for liability rules. This paper identifies these challenges and evaluates how liability rules should be adapted in response. The paper discusses the gaps in liability that arise when AI systems are unpredictable or act (semi)-autonomously. It considers the problems in proving fault and causality when errors in AI systems are difficult to foresee for producers, and monitoring duties of users are difficult to define. From an economic perspective, the paper considers what liability rules would minimise costs of harm related to AI. Based on the analysis of risks and optimal liability rules, the paper evaluates the recently published EU proposals for a Product Liability Directive and for an AI Liability Directive.

1. Introduction

Artificial Intelligence (AI) systems promise to improve societal well-being and increase efficiency in numerous fields, including healthcare, transport, and consumer products. AI systems allow for new approaches to problem-solving, creating the potential for better decision-making. A possible downside of AI is that if mistakes occur in such systems, these errors may be less predictable to humans. The latter may have less control over the functioning of AI, particularly if these systems act (semi-)autonomously. When AI applications possess the characteristics of unpredictability and autonomy, they prove tricky to integrate within the existing liability framework.

This paper identifies the challenges of AI for liability and assesses how liability rules should be adapted to address these issues. It focuses on allocating the responsibility for AI-related harm between producers and users of AI. In particular, the paper considers what constitutes fault if AI actions cannot be reasonably anticipated, how to prove causality if there is no predictable line between AI design and harm, and how to define the responsibilities of producers and users if AI systems act (semi-)autonomously. The paper tackles these questions against the European Commission proposals published

☆ This paper is based on the report "EU liability rules for the age of Artificial Intelligence" prepared in March 2021 by the authors for the Centre on Regulation in Europe (CERRE <http://www.cerre.eu/>), which received the support and/or input of BEUC, the European Commission (DG JUST, DG CNECT and DG GROW), Google, and Vodafone. Martin Peitz gratefully acknowledges support from Deutsche Forschungsgemeinschaft (DFG) through CRC TR 224 (project B05).

* Corresponding author.
E-mail address: miriam.buiten@unisg.ch (M. Buiten).

https://doi.org/10.1016/j.clsr.2023.105794
0267-3649/

in September 2022 to review the Product Liability Directive and to introduce specific EU liability rules for users of AI systems. While the proposed review of the Product Liability Directive[1] aims to clarify product liability concepts for the digital economy, the proposed AI Liability Directive[2] seeks to stipulate specific procedural rules for claims involving AI systems.

The paper considers if these EU proposals strike the right balance when defining who is liable, what standard of liability applies, and how much EU harmonisation they entail. It also reviews the need for AI-specific non-contractual liability rules, taking into account the interaction of such rules with other regulatory instruments. The normative framework of the paper is based on an economic analysis of the law, which involves a contemplation of how liability rules affect the incentives of producers, users, and other harmed parties.

The paper proceeds as follows. Section 2 lays out the broader regulatory framework, which involves the relevant safety regulation complementing liability rules, the general liability standards that apply in EU Member States, and the EU proposals for producers and AI-user liability. Section 3 identifies the gaps in liability resulting from AI's unique characteristics. Section 4 analyses efficient liability rules for AI, discussing what parties should be held liable and what standard of liability is efficient. Section 5 discusses in more detail what changes are appropriate to the EU Product Liability Directive and evaluates the need for an EU liability regime for AI operators. Finally, Section 6 concludes.

2. Existing EU regulatory framework for safety and liability

The rules for non-contractual liability are part of a broader legal framework, including contractual liability rules, general safety rules, as well as sector-specific liability and safety rules that apply in high-risk sectors. The regulatory framework around AI systems is relevant to liability, as it may mitigate risks and help clarify the duties of care of those involved. General or sector-specific EU safety regulations will likely cover many AI systems, and the proposed AI Act and Cybersecurity Act stipulate concrete obligations to mitigate the risks of AI systems.

2.1. *Safety regulation*

Within the EU, the product safety rules, which are primarily ex-ante, are divided into two levels of legislation. On the one hand, some specific laws regulate certain sectors or products, as discussed below. On the other hand, in the absence of such specific requirements, the general rules set out in the General Product Safety Directive (henceforth, "GPSD")[3] apply. The directive aims to ensure that manufacturers place only safe consumer products on the market.[4] In particular, the GPSD obliges producers and distributors to provide safe products to consumers, to take all possible steps to identify any hazards of their products, as well as inform consumers of the existence of such dangers, and, if necessary, to remove dangerous products from the market.[5] After a review of the GPSD, the Commission suggested converting the GPSD into a Regulation, publishing a proposal for a General Product Safety Regulation (henceforth, "GPSR").[6] The proposed GPSR aims to regulate safety risks emerging from new technologies.[7] Additional to the GPSD or soon-to-be GPSR, a Market Surveillance Regulation sets out requirements for accreditation and market surveillance.[8]

With regard to cybersecurity risks, the Cybersecurity Act[9] introduced a uniform European certification framework for ICT products, services and processes. These need to be certified and are assigned a security level. This certification scheme should reduce the risk of cybersecurity vulnerabilities and is therefore an important complement to the liability framework. Additionally, the proposed Cyber Resilience Act[10] would introduce EU-wide cybersecurity requirements for products with digital elements, throughout their whole lifecycle. The proposal seeks to set an EU minimum cybersecurity standard for developing software and hardware products, with specific obligations for different actors within the supply chain. Manufacturers would be subject to cybersecurity risk assessments, conformity assessment procedures and disclosure obligations. The proposal requires manufacturers to ensure that their product has a CE marking, to conduct vendor due diligence if they use third-party parts, and to document their actions.

In an AI-specific context, the proposed AI Act[11] intends to introduce a regulatory framework for AI, instituting obliga-

[1] Proposal for a Directive Of The European Parliament And Of The Council on liability for defective products ("PLD Proposal"), COM(2022) 495.

[2] Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM(2022) 496.

[3] Directive 2001/95 of the European Parliament and of the Council of 3 December 2001 on general product safety [2002] OJ L 11 (General Product Safety Directive).

[4] General Product Safety Directive, art 3.

[5] General Product Safety Directive, art 3.

[6] Commission, 'Proposal for a Regulation of the European Parliament and of the Council on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council, and repealing Council Directive 87/357/EEC and Directive 2001/95/EC of the European Parliament and of the Council' COM (2021) 346.

[7] For an overview see Nikolina Šajn, 'General product safety regulation - European Parliamentary Research Service Briefing' (PE 698.028, September 2021) <https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698028/EPRS_BRI(2021)698028_EN.pdf> accessed on 16 December 2022.

[8] Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011 [2019] OJ L 169/1.

[9] Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act) OJ L 151.

[10] Proposal for a Regulation of the European Parliament and of the Council on horizontal cybersecurity requirements for products with digital elements and amending Regulation (EU) 2019/1020 COM(2022) 454.

[11] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules On Artificial Intelligence

tions for developing and using "high-risk"[12] AI systems and banning[13] specific harmful AI systems. Particularly, high-risk AI systems would be subject to conformity assessments.[14] The definition of AI in the proposed AI Act is purposefully broad, but the exact scope is still being debated.

For high-risk sectors, the EU product safety framework complements the horizontal rules with sector-specific rules. For instance, in healthcare, the Medical Devices Regulation[15] and the In Vitro Diagnostic Medical Devices Regulation[16] apply, while in the domain of transportation, the General Vehicles Safety Regulation,[17] the Approval and Market Surveillance of Vehicles Regulation[18] and the Motor vehicles Insurance Directive pertain.[19] For machinery, the Machinery Directive[20] applies, and a CE Declaration is required.

(Artificial Intelligence Act) And Amending Certain Union Legislative Acts COM/2021/206 final.

[12] Articles 6 et seq. proposed AI Act: AI practices are to be classified as high-risk either by being part of a product required to undergo third-party conformity assessments covered by Union harmonisation legislation listed in Annex II or if the area in which AI is applied is considered risky, as listed in Annex III of the proposal, According to Article 7 proposed AI Act, the Commission is empowered to add high-risk AI systems,

[13] Article 5 proposed AI Act: Any AI practice that used for the purposes of manipulating individuals, exploiting vulnerabilities of a specific group of persons, social scoring, and real-time biometric identification systems in publicly accessible spaces are prohibited. Latter might be allowed under certain conditions if specific objectives are fulfilled.

[14] Article 43 proposed AI Act.

[15] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC [2017] OJ L 117/1.

[16] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU [2017] OJ L 117/176.

[17] Regulation (EU) 2019/2144 of the European Parliament and of the Council on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users, amending Regulation (EU) 2018/858 of the European Parliament and of the Council [2019] OJ L 325/1, art 3(22): "a motor vehicle that has been designed and constructed to move autonomously without any driver supervision".

[18] Regulation 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles [2018] OJ L 151/1. This Regulation took force on 1 September 2020 and repeals Directive 2007/46 of the European Parliament and of the Council of 5 September 2007 establishing a framework for the approval of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles (Framework Directive) [2007] OJ L 263/1.

[19] Directive 2009/103 of the European Parliament and of the Council of 16 September 2009 relating to insurance against civil liability in respect of the use of motor vehicles, and the enforcement of the obligation to insure against such liability [2009] OJ L 263/11.

[20] Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast) [2006] OJ L 157/23.

2.2. *Liability rules*

2.2.1. *Standards of liability*

In most Member States, fault-based liability is the general standard.[21] When claiming damages under a fault-based liability regime, claimants need to prove that the tortfeasor was at fault, that they suffered harm, and that there is a link of causality between the harmful activity and the damage. However, attesting these conditions might prove challenging for harmed parties, which is why other – more claimant-friendly – liability regimes have been introduced for specific situations where the legislator (or case law) aims to grant easier access to damages.

One way of facilitating the filing of damage claims is to hold on to the fault-based liability regime while reversing the burden of proof. A rebuttable presumption of fault or causality can help claimants obtain compensation and reduce information asymmetries between the harmed party and the tortfeasor. A presumption regime may be linked to a diverse set of factual situations generating different types of risks and damages, such as the responsibility of parents for damages caused by their children, employers for employees acting on their behalf, owners of buildings, or persons carrying out dangerous activities.[22]

Another way to make a liability regime more claimant-friendly is to move away from fault-based liability rules, changing the conditions to be attested by the claimant. Under a strict liability regime, claimants are solely required to prove the *default* or the risks taken by the wrongdoers, which are easier to prove than the tortfeasor's intention or negligence.[23] When the likelihood of damage is linked to the unpredictable behaviour of specific risk groups, introducing a strict liability standard can be justified.[24] In such cases – as can be seen with liability rules for autonomous third parties such as animals or children – liability is attributed to the individual responsible for supervising the third party, as they are deemed best capable of adopting measures that prevent or at least reduce the risk. Strict liability may also be justified because the risk of

[21] Konrad Zweigert and Hein Kötz, *Einführung in die Rechtsvergleichung auf dem Gebiete des Privatrechts* (3rd ed, Mohr Siebeck 1996) 650.

[22] Commission, 'Report of 7 May 2018 on the Application of the Council Directive on the approximation of the laws, regulations, and administrative provisions of the Member States concerning liability for defective products (85/374/EEC)' COM (2018) 246, 5-6.

[23] Some forms of strict liability may go even a step further by linking liability simply to the materialisation of risk or making the discharge of liability either impossible or possible only under the proof that the damaging event was caused by an exceptional or unforeseen circumstance that could not be avoided. In effect, those stricter regimes establish non-rebuttable presumptions of a causality link to facilitate the compensation of the victim of damages in situations where the legislator considers it too burdensome or unbalanced to require the victim to prove such causality link.

[24] Commission, 'Liability for emerging digital technologies – Accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial intelligence for Europe' (Staff Working Document) COM (2018) 237, 8

damages is linked to dangerous activities. Some jurisdictions may attribute liability to the person carrying such a dangerous activity (e.g. the operator of a nuclear power plant or an aircraft or the driver of a car) or are ultimately responsible for the hazardous activity (e.g. the owner of a vehicle). The rationale typically is that this person has created a risk, which materialises in damage and, at the same time, also derives an economic benefit from this activity. Those strict liability regimes can apply to diverse factual situations generating different risks and damages, such as the liability of the owners of animals for the damages caused by the animals, the liability of persons carrying out a specified dangerous activity, or damages caused by someone executing a task in the interest of someone else (employee/employer) or by an object that is under their custody.

As a strict liability regime tilts the balance in favour of claimants at the expense of the person responsible, it is generally accompanied by limiting principles, especially regarding the type and amount of recoverable damages. Claimants seeking compensation for more damages than the ones covered by strict liability would need to launch a complementary action against the person responsible under fault-based liability. Strict liability regimes may also be coupled with mandatory insurance requirements reducing the risk for claimants of not being compensated.

2.2.2. *The Product Liability Directive and its proposed revision*

The EU Product Liability Directive (henceforth, "PLD''), adopted in 1985, established a strict liability regime where producers are liable for their defective products regardless of whether the defect is their fault.[25] The PLD is a technology-neutral instrument that fully harmonises product liability rules throughout the EU. It applies to any product sold in the EEA with a three-year limit to recover damages. The PLD assigns liability to the "producer'' (Article1 PLD), which includes the manufacturer of a finished product, the producer of any raw material or the manufacturer of a part, and any person who, by putting his name, trademark or other distinguishing feature on the product presents himself as to its producer.[26]

Over the last few years, the European Commission identified several problems with applying the provisions of the PLD in the context of digital, connected and autonomous systems. One challenge is the complicated product or service value chain, with interdependencies between suppliers, manufacturers and third parties. Another revolves around the uncertainty concerning the legal nature of digital goods, i.e. whether they are products or services. Technologies with autonomous capabilities introduce specific problems for product liability.[27] The Expert Group Report and the White Paper on AI both concluded that some critical concepts in the PLD require clarification to deal with emerging digital technologies.[28] The European Parliament has also called on the European Commission "to review the Directive and consider adapting such concepts as 'product' 'damage' and 'defect' as well as adapting the rules governing the burden of proof''.[29]

On 28 September 2022, the European Commission proposed a revised PLD (henceforth, "PLD proposal") to make it fit for the digital age.[1] The PLD proposal broadens the notions of damage, product, defect, and liable party. The proposal extends the scope of "product''. Under current European Court of Justice case law, the PLD applies to tangible goods.[30] The PLD proposal explicitly includes intangible software and digital manufacturing files as products.[31]

Producer liability under the PLD arises in case of a defect. The PLD defines a "defective'' product as a product that does not provide the safety the consumer is entitled to expect, considering all circumstances. This includes, for instance, the presentation of the product, the use to which it could reasonably be expected that the product would be put, and the time when it was put into circulation. A defect shall be assessed considering "the time when the product was put into circulation.''[32] This concept has raised interpretative questions.[33] The PLD proposal expands the non-exhaustive list of circumstances used to assess the product's defectiveness. It now reads: "(a) the presentation of the product, including the instructions for installation, use and maintenance; (b) the reasonably foreseeable use and misuse of the product; (c) the effect on the product of any ability to continue to learn after deployment; (d) the effect on the product of other products that can reasonably be expected to be used together with the product; (e) the moment in time when the product was placed on the market or put into service or, where the manufacturer retains control over the product after that moment, the moment in time when the product left the control of the manufacturer.''[34]

The PLD proposal extends the liable party from producer to the economic operator, which is defined as "the manufacturer of a product or component, the provider of a related service, the authorised representative, the importer, the fulfilment service provider or the distributor.''[35]

---

[25] Council Directive 85/374/EEC on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [1985] OJ L 210/29 (Product Liability Directive). Given that liability is restricted to defective products, some argue that this is in fact a fault-based liability regime. See e.g. Herbert Zech, 'Liability for AI: public policy considerations' (2021) 22 ERA Forum 147

[26] Article3(2) PLD.

[27] Commission, 'Building a European data economy' (Communication) COM (2017) 09.

[28] COM (2020) 65 final; Commission, 'Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics' COM (2020) 64 final; Expert Group on Liability and New Technologies New Technologies Formation, Liability For Artificial Intelligence And Other Emerging Digital Technologies (2019) (henceforth, "Expert Group Report"), 27-28.

[29] European Parliament, *Resolution on automated decision-making processes: ensuring consumer protection and free movement of goods and services* (2019/2915(RSP)) (EP Resolution).

[30] The PLD applies to "movables" (Art. 2), which the European Court of Justice (ECJ) has interpreted as tangible goods. The ECJ has indicated that the PLD applies to products used while providing a service, Case C-203/99 *Veedfald*, ECLI:EU:C:2001:258.

[31] Article4 (1) and Rec. 12 PLD proposal. Recital (13), however, excludes non-commercially used, free and open-source software from the PLD's scope.

[32] Article 6(1)(b) PLD.

[33] See e.g. Case C-127/04 *Declan O'Byrne v. Sanofi Pasteur MSD* ECLI:EU:C:2006:93.

[34] Article 6(1) PLD proposal.

[35] Article 7 PLD proposal.

The PLD defines damage as death, personal injury, or damage to the product or other property. The notion of defect focuses on consumers' safety expectations to physical harm, excluding possible privacy harm, cybersecurity flaws, or other risks that may arise for IoT products. The PLD proposal adds the "loss or corruption of data that is not used exclusively for professional purposes" as compensable categories of damage.[36]

2.2.3. *Proposed AI Liability Directive*

The proposed AI Liability Directive (henceforth, "proposed AI Liability Directive") will lay down uniform rules around the civil liability of owners and users of AI. The proposal complements the PLD and follows the definition of high-risk in the proposed AI Act. It foresees rules on the claimant's access to evidence of the defendant, allowing (potential) claimants to request access to relevant evidence about a specific high-risk AI system suspected of having caused damage.[37] National courts will oversee and order the defendant's disclosure and preservation of evidence. If a defendant fails to comply with court orders relating to the handling of evidence, a presumption of non-compliance with duties of care is presumed. The defendant may rebut that presumption by submitting evidence to the contrary.[38]

The proposal, moreover, introduces a rebuttable presumption of a causal link between the defendant's fault and the output (or lack thereof) produced by the AI system.[39] For the presumption to apply, three conditions need to be met: (1) proof of fault of the defendant by the claimant, (2) reasonable likeliness that the fault influenced AI's output/failure, and (3) proof by the claimant that AI's output/failure gave rise to damage. For limited-risk AI systems, as defined in the proposed AI Act, the presumption of causality only applies if a national court considers it excessively difficult for the claimant to prove the causal link.[40]

Aside from this exception, the proposed AI Liability Directive, in principle, applies to high-risk AI systems. The proposed AI Act declares an AI system as high-risk based on two conditions: Either the AI system is intended to be used as a safety component of a product, or is itself a product covered by the Union harmonisation legislation listed in Annex II of the proposed AI Act, or the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment according to the Union harmonisation legislation listed in Annex II.[41]

3. Identifying gaps in liability rules for AI systems

To identify gaps in the EU liability rules for claims involving AI systems, we need to assess how the characteristics of AI affect liability claims. Without repeating the extensive literature on the nature of AI, this section highlights critical aspects of AI and their implications for risk, types of harm that may occur, and proving liability.

3.1. *Risks associated with AI*

AI is an umbrella term for various technologies that rely on algorithms, which have different features and are designed for diverse fields of application. AI has been defined in terms of its perceived intelligence,[42] its ability to act autonomously[43] and its characteristic of evolving in an unforeseeable way.[44] From a technical perspective, it is clear that not all algorithms constitute AI. Still, there is no consensus over what subset of algorithms, such as machine learning algorithms or neural networks, is AI.[45] The proposed AI Act defines AI as software developed with machine learning, logic-based and statistical approaches that can generate output influencing the environments with which they interact.[46] The EU AI High-Level Expert Group Report recognises complexity, opacity, openness, autonomy, predictability, data-drivenness, and vulnerability as unique characteristics of AI.[47] When identifying the appropriate liability regime for AI-based technologies, it is relevant to how these characteristics affect the risks of AI.

First, AI can be *complex* because of the involvement of multiple stakeholders and the interdependence of AI components. The various parts of digital goods, such as hardware and digital content, may be sold separately and produced by multiple parties. This can make it difficult to trace the source of a malfunction or attribute liability for the malfunction to a single manufacturer. Injured parties may be confronted by hardware manufacturers, software designers, software developers, facility owners, or others.[48] Consumers may have difficulty prov-

[36] Article 4(6)(c) PLD proposal.
[37] Article 3 proposed AI Liability Directive.
[38] Article 3(5) proposed AI Liability Directive.
[39] Article 4(5) proposed AI Liability Directive.
[40] Article 4(5) proposed AI Liability Directive.
[41] Article 6 proposed AI Act.

[42] John J. McCarthy, Marvin L. Minsky, and Nathaniel Rochester, *Artificial Intelligence* (Research Laboratory of Electronics at the Massachusetts Institute of Technology 1959).
[43] Matthew U. Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, And Strategies' (2016) 29 Harvard Journal of Law & Technology 353, 363.
[44] Urs Gasser and Virgilio A. F. Almeida, 'A Layered Model for AI Governance' (2017) 21 IEEE Internet Computing 58.
[45] See e.g. Miriam C. Buiten, 'Towards intelligent regulation of Artificial Intelligence' (2019) 10 European Journal of Risk Regulation 41; Philipp Hacker, 'Europäische und nationale Regulierung von Künstlicher Intelligenz' (2020) NJW 2142; Scherer (n 46), 359. See also John McCarthy, 'What is Artificial Intelligence? (12 November 2007) <www-formal.stanford.edu/jmc/whatisai.pdf> accessed 13.12.2022.
[46] Commission, 'Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts' COM (2021) 206 ("Proposed AI Act"), art 3 (1): In particular, Annex 1 names three specific techniques that are to fall under the term AI: 'machine learning approaches', including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; 'logic- and knowledge-based approaches', including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; and, 'statistical approaches' that include bayesian estimation, search and optimisation methods.
[47] Expert Group Report (n 29), 7.
[48] Martin Ebers, 'Regulating AI and Robotics' in Martin Ebers and Susanna Navas (eds), *Algorithms and Law* (Cambridge University

ing why their product does not work.[49] They may not have a tangible product at all – AI-based technologies and, more broadly, digital technologies may be offered as a service instead.[50] Regardless of whether the AI system is offered as a product or as a service, multiple parties may be involved in producing or providing it.

The involvement of multiple stakeholders is neither new nor limited to AI systems. Various stakeholders are involved in producing many products, such as cars, which are effectively regulated by the EU's existing liability regime. Nevertheless, having multiple parties involved in making or providing an AI system requires responsibilities to be clearly allocated amongst them. In a litigation setting, the question of which party was responsible for ensuring safety and compatibility in the specific case may arise. This may be the case if, for instance, an AI application makes a mistake because it was not trained on sufficiently rich data.

A related issue is that risks may be correlated for data-driven, probabilistic AI systems. The risk of harm related to AI can increase if interdependent components of AI systems from different manufacturers raise compatibility issues.[51] The tangible devices, such as sensors or hardware, interact with the software components and applications, the data itself, the data services, and the connectivity features.[52] AI systems could become more vulnerable as these systems become interconnected; therefore, the risk of unanticipated or cascading problems grows.

Second, AI systems can be *opaque*. Such a lack of transparency[53] can make it challenging to identify causality because it may be unclear how input resulted in output.[54] Injured parties might not realise that they have been harmed or be unable to trace back the source of the harm.[55] Opaque systems can therefore make it more difficult to hold decision-makers accountable or liable for the outcomes of these systems.[56]

Third, as AI systems become increasingly *autonomous*, it also becomes difficult to trace back outcomes to human decisions and attribute responsibility to a specific actor.[57] Autonomy shifts control away from users and possibly from manufacturers. AI systems that can act independently may be *unpredictable*.[58] The functioning of an autonomous AI system is not necessarily understandable and predictable in the same way as traditional engineering systems.[59] If manufacturers cannot foresee how an AI application will decide or act once placed on the market, it may be difficult to hold them responsible if the AI causes harm.[60] At the same time, if AI systems operate autonomously, users have less control over how these systems function – for instance, robotic vacuum cleaners are supposed to work independently and free up consumers' time.

A system's level of autonomy should be distinguished from its level of automation. A system could be automated, but its output is completely pre-programmed. An application with a high level of automation means that little human supervision is required, but not that the outcomes are unpredictable. Automation is nothing new and is already widely used without challenges to the liability system. For instance, alarm systems are usually automated, or the closing and opening of railway crossings.

What is new and salient about AI-based technologies is that they have the potential to produce an output that is not predetermined by the input. This constitutes autonomous decision-making. As machine-learning and deep-learning capabilities advance, AI systems may be technically able to make predictions independently.[61] AI systems may act in ways that humans would not have considered, reducing the control humans have over the outcomes. An example is C-Path, a machine-learning program for detecting cancer, which found indicators for diagnosing breast cancer that contradicted predominant medical thinking.[62] The ability of AI systems to

Press 2020), 44. See also Christiane Wendehorst, Sale of goods and supply of digital content – two worlds apart?: Why the law on sale of goods needs to respond better to the challenges of the digital age (*Study for the Juri Committee by the Directorate General for Internal Policies, Policy Department C*, 2016) <https://www.europarl.europa.eu/cmsdata/98774/pe%20556%20928%20EN_final.pdf> accessed on accessed on 16 December 2022, 7.; BEUC (The European Consumer Organisation), 'PRODUCT LIABILITY 2.0 How to make EU rules fit for consumers in the digital age' (BEUC-X-2020-024, 07 May 2020) <www.beuc.eu/publications/beuc-x-2020-024_product_liability_position_paper.pdf> accessed on 07 August 2022, 18-19; Yaniv Benhamou and Justine Ferland, 'Artificial Intelligence & Damages: Assessing Liability And Calculating The Damages' in Giuseppina (Pina) D'Agostino, Carole Piovesan and Aviv Gaon (eds.), *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law* (Thomson Reuters Canada 2020) 6.

[49] Ebers (n 51), 44.

[50] See further Section 5 below.

[51] A possible response by manufacturers is to exclude warranties when third-party software is used. However, to encourage competition in the market, it may be beneficial to find a legal solution that encourages manufacturers to allow third-party developers access to their products.

[52] Benhamou and Ferland (n 51) 6, referring to COM (2018) 237.

[53] Commission, 'On Artificial Intelligence – A. European approach to excellence and trust' (White Paper) COM (2020) 65 final, 17. This is also referred to as the "black box-effect" of AI (see COM (2020) 65 final 1)

[54] Ebers (n 51), 48.

[55] Expert Group Report (n 29), 33.

[56] Lilian Edwards and Michael Veale, 'Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for' (2017) 16 Duke Law & Technology Review 18.

[57] EP Resolution, 6, number 7

[58] Scherer (n 46), 363-64; Peter M. Asaro, 'The Liability Problem for Autonomous Artificial Agents' (2016) AAAI Spring Symposia 2; Harry Surden and Mary-Anne Williams, 'Technological opacity, predictability, and self-driving cars' (2016) 38 Cardozo Law Review 121.

[59] Asaro (n 61), 2.

[60] Hannah R. Sullivan and Scott J. Schweikart, 'Are current tort liability doctrines adequate for addressing injury caused by AI?' (2019) 21 AMA journal of ethics 160; Asaro (n 61), 2.

[61] Daniel Schönberger, 'Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications' (2019) 27 International Journal of Law and Information Technology 171, 193.

[62] Scherer (n 46), 363-364. A number of AI tools are available to detect health conditions. For cancer image search, see Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, et al. 'Human-centered tools for coping with imperfect algorithms during medical decision-making' (2019) Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. For cancer image classification, see Amirhossein Kiani,

come up with new solutions is amongst its great benefits. If AI systems are, on average safer than their non-AI counterparts,[63] there are opportunity costs of not relying on AI. If an AI system is not employed, a less safe human alternative will take its place. Uptake of such technology should be encouraged, but new risks and challenges for liability should be addressed.

Nevertheless, the more autonomy AI systems achieve, the more tenuous it becomes to attribute legal responsibility for their actions to human beings.[64] Fully autonomous AI presents different challenges for liability than AI that requires some level of human supervision. If a task is fully delegated to AI, humans need to be able to rely on it functioning on its own. "Monitoring" the AI system would then likely entail periodic check-ups of the output. What is needed in terms of monitoring or oversight would depend on the type of system – for instance, a chatbot would need a different kind of monitoring than an AI lawn mower. Systems evolving over time, such as a hiring algorithm or software to reduce car emissions, would need regular performance checks. Overall, monitoring an AI system would not be that different from monitoring a traditional system and would depend on context. It is, however, clear that fully autonomous AI systems would not require permanently observing the system as it operates – this would defeat the purpose of an autonomous AI system.

Monitoring takes a different form if AI supports decision-making,[65] but humans still need to oversee the system. For instance, drivers of semi-autonomous cars still need to monitor the road constantly. When AI requires human-machine interaction, the question arises in which circumstances users can rely on the AI system and at what point they should override its decisions. Consider, for example, the responsibility of a physician relying on a medical AI application to decide on treatment. If the AI tool makes an error, the patient may suffer harm due to an inappropriate drug recommendation. The question is under what circumstances the physician should rely on the decision of the AI application and when she would be held liable for following an incorrect recommendation of the AI application.[66] The liability question may be more straightforward when there is no human-machine interaction and the AI system functions fully autonomously.

3.2. *Gaps in liability associated with characteristics of AI*

AI that possesses the characteristics of autonomy, unpredictability, complexity or opacity introduces new risks and challenges to attributing responsibility. Based on these challenges, three possible gaps in the existing liability regime can be identified. First, it may be unclear what constitutes fault, or a product defect, if AI actions cannot be reasonably anticipated. This raises questions about how to divide responsibility between producers and users. Second, it is difficult to prove causality if there is no traceable and predictable line between AI design and harm. Third, the types of damages caused by AI may not be included in the recognised categories of harm for recovery under a liability rule.

First, *what constitutes fault* is unclear if AI actions cannot be reasonably anticipated.[67] Establishing responsibility is quickly done when operators use an AI system to cause harm deliberately, but it is much more difficult for unintended harm.[68]As a result of the unpredictability of AI systems and the lack of control on the side of users, complex automated systems pose unique problems to fault-based liability regimes. Fault on the side of users would need to be established in terms of a failure to maintain the automated system or to oversee its functioning. Starting from the premise that AI systems are primarily tools, fault-based liability can continue to hold their users to a duty of reasonable care while using it.[69] However, it is not clear whether the decision of a user to put an autonomous system into operation could be considered negligent if the system causes harm, at least not in all Member States.

Fault-based liability may run into problems, particularly for decision-assistance AI, designed to interfere with human decision-making. If AI systems are to improve upon human decision-making, and we lack understanding of how it does this, it is questionable whether humans be considered negligent for relying on the AI system when this leads to harm. The example above from healthcare illustrates this: A patient attempting to collect damages from a physician or a hospital may have difficulty proving that either of them was at fault when they relied on AI. It may also be unfair and inefficient to allocate full responsibility to healthcare providers if an AI system makes a harmful recommendation or decision. This would disconnect accountability from the locus of control.[70] To establish if a physician was at fault when an AI system was

Bora Uyumazturk, Pranav Rajpurkar, Alex Wang, Rebecca Gao, Erik Jones, Yifan Yu, Curtis P Langlotz, Robyn L Ball, Thomas J Montine, et al. 'Impact of a deep learning assistant on the histopathologic classification of liver cancer' (2020) 3 NPJ digital medicine 1.

[63] See e.g. UK Office for Product Safety & Standards, Study on the Impact of Artificial Intelligence on Product Safety, tps://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1077630/impact-of-ai-on-product-safety.pdf.

[64] Mark A. Chinen, 'The co-evolution of autonomous machines and legal responsibility' (2016) 20 Virginia Journal of Law & Technology 338.

[65] A large number of use cases and corresponding references are provided by Vivian Lai, Chacha Chen, Q. Vera Liao, Alison Smith-Renner, Chenhao Tan, 'Towards a science of human-AI decision making: A survey of empirical studies' (2021) https://arxiv.org/abs/2112.11471, accessed on 16 December 2022.

[66] See further H. Smith & K. Fotheringham, 'Artificial intelligence in clinical decision-making: Rethinking liability' (2020) Medical Law International, 20(2), 131.

[67] William D. Smart, Cindy M. Grimm, and Woodrow Hartzog, 'An education theory of fault for autonomous systems' (2021) 2 Notre Dame Journal on Emerging Technologies 34; Jin Yoshikawa, 'Sharing the costs of artificial intelligence: Universal no-fault social insurance for personal injuries' (2019) 21 Vanderbilt Journal of Entertainment & Technology Law 1155 referring to Curtis E. A. Karnow, 'The application of traditional tort theory to embodied machine intelligence' in Ryan Calo, Michael A. Froomkin, and Ian Kerr (eds), *Robot Law* (Edward Elgar Publishing 2016), 52. See also Andrew D. Selbst, 'Negligence and AI's Human Users' (2020) 100 Boston University Law Review 1315, 1331 et seq.

[68] See also Von Ungern-Sternberg (n 157) 6; Weston Kowert, 'The foreseeability of human-artificial intelligence interactions' (2017) 96 Texas Law Review 181, 191.

[69] See for a US perspective Andrew Selbst, 'Negligence and AI's human users' (2020), Boston University Law Review 100, 1315, 1320.

[70] Smith and Fotheringham (n 69).

involved, one must ask if it was reasonable to rely on the system in the given situation. If the physician depended on a certified, broadly used AI tool, the error was not obvious, and she followed safety standards and best practices, she is likely not at fault for relying on it.[71]

These aspects raise questions about the division of responsibility between manufacturers and users. It may be unclear if the harm is the result of a product defect or improper use if it is at least partly attributed to the general unpredictability of the system. Autonomous systems will likely shift responsibility towards manufacturers.[72] The question arises what the limit of producer liability is for AI systems with a high level of autonomy – for instance, if any harmful action constitutes a defect or if we accept that well-functioning AI systems may nevertheless cause harm from time to time. It is not necessarily clear what liability should continue to fall on owners and users. Moreover, problems may arise when dividing responsibility amongst manufacturers and other stakeholders involved in the product's functionality, such as data providers.

Second, proving *causality* can also be problematic if there is no traceable and predictable line between AI design and harm. Developers do not control automated systems quite the same way that, for instance, car manufacturers manage how airbags deploy.[73] A related issue is what level of safety can be expected of AI systems that are supposed to take decisions or actions autonomously. In terms of producer liability, this raises the question of what constitutes a "defect".

Third, depending on how an AI-based technology is consumed, AI can be prone to different *types of harm* in case of an accident than traditional products. Some of these categories of harm may not be recognised in liability law, particularly in producer liability. If an AI system controls a physical, tangible system, it can create physical harm in the same way as traditional products. For instance, an autonomous vehicle, surgical robot or robot vacuum cleaner can cause physical damage much like their traditional counterparts can. Even if an AI-based technology is consumed as a service, its decisions have consequences in the physical world, and harm may occur in the familiar categories. For instance, if a hospital uses an AI-based medical diagnostics tool for which they pay a monthly subscription fee, the consequences may still be physical harm if the tool makes an error.

Beyond this, AI tools – or, more broadly, digital tools – may create types of harm that are less prominent or relevant for traditional products. For example, an AI tool interconnected with a company's logistical data could cause damage to this data, disrupting the company's operations. An AI-based app vulnerable to cyberattacks could cause financial and privacy harm to its users. A security breach on a robot vacuum could result in a map of the user's home being shared with others or third parties getting access to the local network of the user.[74] An AI tool that takes hiring decisions, for instance, can cause harm by taking discriminatory decisions, or an AI chatbot could produce hate speech or commit copyright violations. Given that AI may make certain types of harm more likely, it may be relevant to reconsider the harms recognised in liability law, particularly producer liability.

[71] Schönberger (n 64) 197.

[72] Selbst (n 70) 1322.

[73] Smart, Grimm, and Woodrow (n 70) 12-13.

[74] F. Ullrich, J. Classen, J. Eger, & M. Hollick, 'Vacuums in the cloud: analyzing security in a hardened IoT ecosystem' (2019), Proceedings of the 13th USENIX Conference on Offensive Technologies, 7.

4. Efficient liability rules for AI

Imposing tort liability on those engaged in activities that may cause harm operates as a mechanism for internalising harmful externalities. One objective of tort law is to incentivise potential wrongdoers to invest in safety at an efficient level by making them compensate damages.[75] Investing in safety measures is costly since precautions require time and resources. Therefore, zero risks are typically not the socially optimal level of risk since risk reduction naturally comes at an increasing marginal cost.[76] Depending on the particular application, preventive measures can take different forms. For example, one could deploy additional testing of AI-based solutions, consult outside experts as certifiers, commit to human supervision, or apply a careful design of the human-machine interface to reduce human decision-making errors.

4.1. *Liability standards: fault-based or strict liability*

When choosing between a fault-based or a strict liability regime for AI, relevant points to consider are information costs, the role of the injured party, the value of the (risky) activity and the type of risk.

4.1.1. *Information costs and incentives of the victims*

Under a fault-based regime, the owner of an AI system, for instance, a drone with AI technology, is held liable if they fail to take the safety precautions demanded by the standard of care. The owner is induced to take efficient precautions if lawmakers and courts correctly determine the duty of care. A bar set too high or too low will incentivise the drone owner to take a suboptimal level of precautions. If the drone has AI capabilities, a fault-based regime is potentially suboptimal if courts cannot accurately assign liability. As discussed above, legal conditions for liability, such as fault and causation, may be challenging to prove for AI applications. As a result, the efficient level of precautions may be challenging to determine for AI on a general level. Specifically, the efficient level of precautions may depend on the technical possibilities to control the actions of AI when designing it. There may be a trade-off here between the safety of AI and its sophistication. That is, AI

[75] Robert B. Cooter and Thomas Ulen, *Law and Economics* (6th edn, Pearson 2012), 189-190.

[76] Also, additional measures may be less effective. Assuming that precautions reduce the likelihood of an accident or the amount of harm, but at a decreasing rate of success, the optimal expenditure on precautions is finite. The efficient level of precaution prevails when the additional cost of a precaution measure equals the resulting reduction in expected costs of harm ("marginal costs equal marginal benefits"). Ethical concerns can be raised against such a cost-benefit approach. Based on ethical concerns it is conceivable to prohibit or at least limit the use of AI for certain types of activities.

may offer more benefits to users but may also become increasingly complex or unpredictable. Lack of predictability and autonomy could increase risk. If owners and users cannot control an AI system, fault-based liability does not serve its goal of steering them towards more cautious behaviour. In other contexts, such a lack of control has been a reason to introduce risk-based or strict liability.[77]

The advantage of strict liability is that the legislator or the court does not need information on the optimal level of precaution. A strict liability rule induces the drone owner to take optimal precautions because it shifts all the costs of an accident onto them. Theoretically (under perfect compensation), a strict liability rule internalises the costs of harm by requiring the injurer to pay for the social costs of her activity, regardless of the level of care taken.

However, because the injurer bears all the costs, a strict liability fails to set incentives for victims to take the appropriate care in situations where they, too, can affect the likelihood of an accident. The economics literature defines this as a double moral hazard problem.

4.1.2. *Level of activity and innovation*

Shifting the total costs to the damaging party using a strict liability rule induces them to observe the optimal level of care and the optimal activity level.[78] Suppose an activity is inherently risky, even despite efficient precautions. In that case, we may want to refrain injurers from engaging in this activity altogether - or, at least, to reduce the level of this activity.[79] A fault-based regime does not achieve this outcome since an injuring party can avoid paying for the costs of her activity by taking the required level of care. This explains why most jurisdictions impose strict liability for driving a car, for instance. AI applications could also cause serious harm, even if proper precautions are taken, particularly when AI systems operate in physical space (such as drones or autonomous vehicles). An advantage of strict liability is thus that it could induce AI producers and users to internalise all costs of harm associated with the technology.

However, the flip side is that if an activity benefits society, the potential wrongdoer may become too careful. Strict liability may reduce their activity below the efficient level because negative externalities (i.e. harm) are internalised while positive externalities (i.e. external benefits to society) may not all flow back to them. AI applications produce clear benefits to third parties: cars with autonomous features may be safer, AI diagnostic tools may be superior to humans in detecting diseases, and algorithms produce all types of digital services that consumers enjoy. When employing AI reduces harm or produces benefits compared to the alternative, not using AI will result in opportunity costs.

Moreover, investments in AI applications and their employment may contribute to innovations in AI in other fields as well. A concern for any liability rule and, in particular, strict liability is that start-ups deploying AI may not be able to bear the associated risk and thus go bankrupt, which would shift at least part of the liability to other parties or the injured party if the start-up fails to compensate the incurred damage fully.[80] Furthermore, foreseeing these problems, entrepreneurs may not put their efforts into such a start-up in the first place or may not receive funding. Mandatory insurance could, at least partly, address this issue. However, this would come at the cost of negatively affecting the injurer's incentives to efficiently reduce harm and thus prove to be rather costly.

In this context, it should be acknowledged that liability does not necessarily chill innovation: it may also encourage firms to develop risk-mitigating technologies and improve the design of their products to reduce the likelihood of harm and increase user trust and take-up.[81] Absent liability, there are often insufficient incentives to do so, and potential users may correctly anticipate such a problem and delay adoption. In other words, liability can be a catalyst of innovation.

4.1.3. *Types of risks*

To address AI liability, it is helpful to elaborate on several economic environments in which third parties experience damage, and the effort decision by a party affects the likelihood that damage occurs or the severity of the damage. An important distinction is whether risks are idiosyncratic across people constituting the third party or highly correlated.

In the case of idiosyncratic risk, from an individual perspective, the damage remains a random event, but for the liable party, the outcome is somewhat predictable. For instance, a firm may invest in reducing the fraction of products that pose a risk to third parties. While an accident is highly unpredictable for an individual, a firm faces an average number of accidents, which can be predicted rather well. Fault-based liability may be based on calculating an optimal number of accidents (based on a cost-benefit analysis), and the firm may have to contribute to a pool if the number exceeds of observed accidents is above the optimal number, with a payment that increases in the number of accidents.

Of course, if the fault is directly observable with little cost for the legal system, damage payments due to fault-based li-

---

[77] For instance, parents may be liable for harm caused by their children, or owners of animals may be liable for harm caused by these animals. See Section 2 above.

[78] This does not, however, induce victims to take care.

[79] Gerhard Wagner, 'Robot Liability' in Sebastian Lohsse, Reiner Schulze, and Dirk Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Nomos 2019), 30 notes with respect to liability for AI: "shielding businesses from liability for the harm that they cause, for instance, with a view to fostering innovation, also seems problematic. This is not to say that innovation is unimportant or that incentives to innovate should not be generated. It is doubtful, however, whether the liability system is the preferred tool to create such incentives. To shield certain parties from responsibility for the harm that they actually caused amounts to a subsidization of dangerous activities, leading to an oversupply of such activities."

[80] Liability rules may then act as entry barriers. See also Miriam Buiten, Alexandre de Streel, and Martin Peitz, 'Rethinking liability rules for online hosting platforms' (2020) 28 International Journal of Law and Information Technology 139, 153 with respect to liability rules for online platforms.

[81] Alberto Galasso and Hong Luo, 'When does product liability risk chill innovation? Evidence from medical implants' (2018) NBER Working Paper Series, 25068 <https://www.nber.org/system/files/working_papers/w25068/w25068.pdf> accessed 5 December 2022.

ability can be assessed in individual cases as the third party receives the entire damage in case fault is established. Strict liability would award damages in all cases independent of the level of care. Therefore, claiming damages under a strict liability regime proves more accessible for victims.[82]

In theory, the application of fault-based liability could work as follows: There is an optimal level of protection implying certain damage in case of failure (damage quantified in Euro, say X, and a probability p^* this damage happens). Keeping the size of the damage constant, we can focus on the failure probability. When the optimal (or, in some cases, unavoidable) failure probability is p^*, one can consider that $(p^0 - p^*)X$ is the part of expected harm for which the firm is at fault, when p^0 is the actual failure rate and $p^0 > p^*$. Here, the firm expects to pay damages that correspond to the incremental harm from the failure rate being higher than p^*.[83] In expectations, this is payment to be made. Since a failure occurs with probability p^0, the expected payment is p^0D, which should be equal to $(p^0 - p^*)X$. Once a failure is observed, it is inferred that the firm's fault increased the risk (provided that $p^0 > p^*$) and, therefore, has to pay $\frac{p^0-p^*}{p^0}X$. For example, suppose that the investigation concluded that the optimal failure probability is 1% and the actual failure probability was 5%, then due to fault, with total harm of € 1 million to third parties, damages of € 800k should be awarded as damages. Thus, in theory, the idea of fault can be applied to such probabilistic events.

In terms of incentives (taking their presence in the market as given), fault-based and strict liability regimes perform equally well in theory. From an ex-ante perspective, when deciding the level of care, a firm minimises the sum of expected damages and precaution costs C as a function of $1 - p^0$, which, under strict liability, is $p^0X + C(1 - p^0)$ with respect to p^0. This implies that the firm chooses the risk such that $X = C'(1 - p^0)$. Under fault-based liability, a firm minimises $(p^0 - p^*)X + C(1 - p^0)$ with respect to p^0. This implies that the firm chooses the risk such that again $X = C'(1 - p^0)$.

The above argument shows that fault-based, as well as strict liability, lead to the same (efficient) level of care. However, the application of fault-based liability (as constructed above) leads to practical problems since it requires the court's ability to calculate optimal and actual risk. Strict liability does not suffer from this practical problem, as, in our example, in case of failure, the court would award simply € 1 million. In many instances, including many cases involving AI risk, individual risks of third parties are highly correlated, and failure is a rare event. For example, think of insufficient protection of personal data that is hacked despite an AI system that is supposed to detect such threats. This makes it next to impossible to calculate actual failure rates based on past observations. Fault-based liability must then rely on predefined levels or principles of precaution.[84] Therefore, if the individual risk is highly correlated and a failure is a rare event,[85] practical considerations make strict liability the preferred option, as it does not require the calculation of actual and optimal failure rates.

From a dynamic perspective, a downside of strict liability is that the expected payment is bigger for an innovator than under fault-based liability. The higher likelihood of expensive damage claims might lead to a socially insufficient level of innovation, particularly if the innovator cannot internalise all the social benefits arising from the creation. The potential innovator might then refrain from entering the market or scaling up activity, abstaining from socially valuable innovation.[86]

Compared to a fault-based regime, the advantage of strict liability is that legislators and courts do not need information on the optimal level of precaution in designing and testing AI-based solutions. By shifting the total costs of harm on injurers, a strict liability rule incentivises injurers to reduce their activity level in cases where AI applications are inherently risky, even if sufficient precautions are taken. A disadvantage of strict liability lies in that its application might reduce the beneficial use of AI applications below the efficient level, for instance, if their superior performance reduces harm to society compared to not employing AI. If an individual risk is highly correlated and a failure is a rare event – a possible high-risk environment – practical considerations make strict liability the preferred option.

82 The damaging party, however, might only partially bear the burden of higher damage payments. In particular, when a firm sells a product, it can pass, at least part of, the increased expected cost per unit to its customers. Such pass-on reduces the attractiveness for customers to adopt the particular AI system, which implies an opportunity cost to the damaging party. Therefore, the damaging party may still have strong incentives to reduce the number of accidents.

83 This notion of fault-based liability differs from the notion according to which the firm that does not satisfy a certain level of precaution has to cover all damages. This alternative notion is further discussed in the footnote below.

84 If there is a functional relationship between the level of precaution and risk p^0, then according to the precaution-based notion of fault-based liability, damages X have to be paid if $p^0 > p^*$ and zero otherwise. If the court only requires proof of precautionary measures, such fault-based liability does not require an assessment of failure probabilities. In this case, expected damage payments are p^0X if $p^0 > p^*$, where p^* is the failure rate that follows from the predefined level of precaution (duty of care) and zero if $p^0 \leq p^*$. If a lack of fault is shown by providing proof of precautionary measures, the court will not need to assess failure rates. The latter notion of fault-based liability provides stronger incentives to implement a predefined level of precaution than strict liability when considering small deviations from that level. As seen above, under strict liability the firm chooses the risk according to $X = C'(1 - p^0)$. Under the latter notion of fault-based liability the firm typically will implement p^* since its damage payment jump upward as it move from compliance to non-compliance of the requested level of precaution. To obtain efficient effort in risk reduction then requires that the predefined level precaution is set such that this achieves the optimal level of risk. Under strict liability, there is no need to address the issue of defining a required level of precaution.

85 As is discussed in section 5 below, the regulatory definition of a high-risk AI system may differ from this context of highly correlated risks.

86 The chosen liability regime should therefore be seen in the context of public policy towards innovation. The choice of strict instead of fault-based liability increases the call for public support to innovations to compensate for the higher expected payments to injured parties.

4.2. *Who should be liable*

When multiple parties affect the risk of harm, the question arises of who should be targeted by the liability rule. From a welfare perspective, this should be the least-cost avoider, i.e., the party which can minimise harm at the lowest cost. To the extent that some harm-reducing activities are complementary, this may imply that liability rules should target multiple parties.

As explained above, in many AI-based solutions, several parties are involved in providing a product or service. While the existence of damage may be easy to prove in a court, the question remains as to who should compensate for the damage and to what extent. When examining different liability rules' effects, it is necessary to specify how failures can occur with several parties. We distinguish between two polar environments. In the first environment considered below, care is cumulative; that is, the provision of care by one party is a perfect substitute for care provided by another party. In the second environment, care by all parties is essential; that is, providing care by one party is a perfect complement to care provided by another party. Strict liability says that the total damage has to be compensated. As is often the case, it is unclear which of the parties is to blame. For simplicity, suppose that two parties symmetrically contribute to the risk.

4.2.1. *Substitute care*

First, consider the substitute case. If at least one of the two parties engages in an effort, the risk is assumed to be p^*, while if none of the two exerts effort, the risk is supposed to be p^0. We take that the socially efficient decision is that one of the two parties exerts effort. If in case of an accident, it cannot be verified which party did not exert effort, the total harm is X. One simple rule would be to allocate damages to the parties equally. With strict liability, each party would have to pay $X/2$. Such a rule cannot achieve an efficient level of effort: each party will underinvest in care to reduce risk, anticipating that they will only have to pay half of the damages.[87]

The efficient solution would be if the least-cost provider exerted the effort to reduce the risk of an accident. If this party can be identified at the outset, one may assign liability to this party. However, this may not be easy to do. Alternatively, the law could specify that a certain type of party will be held liable no matter whether their effort cost is lower than that of the other parties. If this party bargains efficiently with the other party, both may agree to shift liability to the least-cost provider. This would guarantee an efficient level of effort at the lowest cost. Similarly, a fault-based liability rule as discussed above assigns damages $\frac{p^0-p^*}{p^0}X$ to one party according to a pre-specified rule. As discussed above, it will be more difficult to implement such a fault-based liability rule in high-risk environments.[88]

4.2.2. *Complement care*

Second, consider the complement case. Here both parties have to exert effort to reduce the risk from p^0 to p^*. We assume that the socially efficient decision is that both parties 1 and 2 exert effort; i.e., the total cost of effort provision satisfies $C_1(1-p^*)+C_2(1-p^*) < (p^0-p^*)X$. Strict liability that allocates the total harm amongst the two parties according to some exogenous sharing rule does not necessarily achieve the efficient effort.[89] If the two parties are symmetric, effort provision by the two parties is efficient if $2C(1-p^*) < p^0X$. If each party has to bear half of the damage, a party exerts effort if $C(1-p^*) < (1/2)(p^0-p^*)X$ provided that it expects the other party to exert effort as well. If both parties behave that way, an efficient effort is provided.

However, if a party is sceptical about the other party's level of effort, it will not exert effort since this does not reduce the probability of an accident. Thus, there may be a coordination failure. Coordination failures can be avoided if parties can provide proof of effort that is verifiable in court and if a party that does not provide proof, will be held fully liable.

If the effort is not binary, but its level can be adjusted, both parties will exert a socially inefficient level of effort. Simply assigning the total damage to the two parties leads them to invest too little in care. To achieve an efficient level of care, the overall payment must be larger than the harm that is inflicted (above the efficient level). The incremental expected payment from not exerting effort must be equal to $(p^0-p^*)X$ for each party; from a legal perspective, this means that there may need to be punitive damages to implement the socially efficient level of care.

The feature of effort being complemented may be identified in particular in high-risk environments because the effort of all parties is needed to keep risk at bay.[90] For example, self-driving cars require reliable sensors and properly functioning

---

[87] When both parties are equally good at reducing risk, this can be seen as follows. The probability of harm depends on the joint cost the two parties incur, $p(C_1+C_2)$ with $p' < 0$ and $p'' > 0$. The welfare-maximizing solution satisfies $p'(C)X+1=0$. If each party has to pay for half the damage, party 1 minimizes $p(C_1+C_2)X/2+C_1$ with respect to C_1 and party 2 minimizes $p(C_1+C_2)X/2+C_2$ with respect to C_2. Thus, parties incur costs C with $p'(C)X/2+1=0$. Hence, the overall level of care is less with this solution than in the welfare-maximizing one.

[88] With the precaution-based notion of fault-based liability, it may be possible to define specific measures of precaution for each party separately. In this case, the party providing an inadequate level of precaution would have to pay the damage in case of a failure. However, it is likely that the requirement of specific levels of precaution for each party will lead to an inefficient outcome as the opportunity cost of individual efforts are unknown upfront and a shift of liability to the least-cost provider is not possible.

[89] When both parties are equally efficient in reducing harm given that the other party has contributed more and damage is shared equally between the two parties, no party has an incentive to invest more in reducing the probability of harm than the other party. The problem for party 1 becomes to minimize $p(\min\{C_1, C_2\})X/2+C_1$. For $C_1 \leq C_2$, this gives $p'(C_1)X/2+1=0$. Thus, the largest effort in harm reduction that can be supported by the behaviour of rational parties satisfies $C_1=C_2=C/2$ with $p'(C/2)X/2+1=0$. By contrast, the welfare-maximizing solution satisfies $p'(C/2)X+1=0$. Hence, under this liability rule both parties spend too little on harm reduction from a welfare point of view.

[90] We acknowledge that complementarity is not specific to AI (see Michael Kremer, 'The O-ring theory of economic development' (1993) 108 The Quarterly Journal of Economics 551). At the root of the Boeing 737 Max crashes lies a malfunctioning sensor and its interaction with a software. More precisely, "erroneous AOA sensor reading triggered the plane's automated Maneuvering Charac-

AI-based software. If only one of the two malfunctions, this is sufficient to increase the probability of harm significantly.

One could ask whether assigning strict liability to one prespecified party can achieve an efficient level of care. The party that is subject to liability may contract with the other party. However, moving part of the liability risk to this other party creates a free-riding problem, as both parties are only subject to part of the liability risk. Thus, assigning liability to one party and efficient contracting cannot resolve the under-provision problem as long as parties only have to cover the harm that has been incurred. A similar issue arises for fault-based liability rules that only account for incremental harm beyond the efficient level. It is thus important to acknowledge that in the presence of complementarities in which individual effort cannot be proved in court, merely compensating damages will lead to an inefficient level of care. This holds even under strict liability.

If each of the parties providing care as perfect complements is fully liable for the total damage, efficient care will be provided. However, the harmed party will then receive double damages. In the spirit of fault-based liability, by assigning damages to each party based on the incremental harm above the efficient level, under some conditions, the total payment can then be kept below the money equivalent of the total damage, and still, the incentives for effort provision are efficient.[91]

To use a numerical example, suppose a lack of care by either one of the two parties implies that an accident occurs with probability $p^0 = 5\%$, while with efficient care by both parties, this probability is reduced to $p^* = 3\%$. Expected incremental harm from a lack of care is 2% times damage X. When X is € 1 million, this amounts to € 20 000. In case of an accident, each party would be required to pay $((p^0 - p^*)/p^0)X =$ €400 000. Thus, the total payment would be €800 000, which is less than the total damage of € 1 million. As discussed above, the difficulty in applying this idea in practice is the lack of information by the court about p^0 and p^*.

Overall, for many AI-based solutions, several parties are involved in providing the product or service. If care by each party is essential to avoid a failure, and courts cannot verify the source of the failure, even strict liability leads to a socially inefficient level of care when no punitive damages are allowed.

4.2.3. *Holding producers and operators liable*

There are several reasons to hold users and owners ("operators") of AI liable next to producers. Liability rules should induce both producers and operators to take an efficient level of care in designing, testing, and employing AI-based solutions. Liability for operators encourages them to take precautions when monitoring AI systems that are not fully autonomous.

For highly autonomous AI systems, liability provides an incentive for operators to keep the system up to date and to ensure that it is used properly. While producers control the product's safety features and provide the interfaces between the product and its operator, the operator exercises control over the use of the system. The operator decides in which circumstances the system is used and is in a position to oversee it in real-world situations. It is, therefore, appropriate to attribute some liability to operators who choose to delegate decisions to AI systems.

Another reason to hold operators liable is that they benefit from the use of AI, so it is appropriate that they bear (some of) the associated costs. Here it should be borne in mind that society can also benefit from AI being used as a safer alternative to non-AI technology. Liability rules should not hamper the uptake of AI technology. So, for instance, if autonomous cars are significantly safer than human-driven cars, the acceptance of these products should be promoted. This is an argument not to introduce a stricter liability for AI systems than for their traditional counterparts.

When designing liability rules, policymakers should recognise that these rules shift costs and, therefore, may influence the design choices of producers in delegating decisions to AI-based systems or humans.

AI systems shift the locus of control away from users towards manufacturers.[92] For technical products that do not rely on AI, the manufacturer controls the product's safety features and provides the interfaces between the product and its user, while the user exercises control over the mechanical device when employing it in real-world situations.[93] For AI systems, users will be able to exert much less control. As a result, accidents will become less dependant on the care taken by the individual user. The liability of the user is likely to increasingly recede into the background, meaning that the role of liability of the manufacturer becomes more significant for injured parties to obtain compensation.[94] In short, where producers are in a better position than consumers to control risk, an incentive-based approach would shift the relative burden of liability towards producers.[95] This incentivises producers to reduce the AI system's risk through designing and manufacturing the system.

While AI systems shift the locus of control to producers, producers do not influence the final use of the AI system. It is, therefore, justified to attribute some liability to the party who owns the AI-powered product (owner) or who uses it (keeper):

teristics Augmentation System (MCAS) anti-stall software" (Benjamin Zhang, 'Boeing and Ethiopian investigators confirm a faulty sensor was triggered on the 737 Max shortly before it crashed' (*Business Insider*, 4 April 2019) <https://www.businessinsider.com/boeing-ethiopian-investigators-confirm-bad-sensor-triggered-faulty-software-before-crash-2019-4?r=US&IR=T> accessed on 5 December 2022)

[91] See Robert B. Cooter and Ariel Porat, 'Total Liability for Excessive Harm' (2007) 36 Journal of Legal Studies 63. From an economics perspective, this is a simple application of the strategic issue in the provision of Cournot complements.

[92] Wagner, 'Robot Liability' (n 82) 37.

[93] Gerhard Wagner, 'Robot, inc.: Personhood for autonomous systems?' (2019) 88 Fordham Law Review 591, 602.

[94] Wagner, 'Robot, inc.: Personhood for autonomous systems?' (n 96); Astrid Seehafer and Joel Kohler, 'Künstliche Intelligenz: Updates für das Produkthaftungsrecht?' (2020) Europäische Zeitschrift für Wirtschaftsrecht 213; Carina Lutter, 'Fragen der Produkthaftung im Hinblick auf den Betrieb unbemannter Schiffe' (2017) 5 Recht der Transportwirtschaft 281

[95] Alberto Galasso and Luo Hong, 'Punishing Robots: issues in the economics of tort liability and innovation in artificial intelligence' (2018) The Economics of Artificial Intelligence: An Agenda, *University of Chicago Press*, 493.

the "operator".[96] Much of today's AI technology is not fully autonomous and requires at least some level of human supervision. Operator liability encourages them to take precautions in supervising the AI system.[97] Even for highly autonomous AI systems, the operator decides if and how to employ them. Liability provides an incentive for operators to keep an AI device updated and ensure that it is used properly.[98] The operator, moreover, benefits from employing AI. Holding producers liable for every case of harm, even those they have no control over and are not capturing the benefits from, may harm innovation.[99]

4.3. *Considerations for efficient AI liability rules*

To summarise, from an economic point of view, tort liability should induce producers and users to take an efficient level of care in designing, testing, and employing AI-based solutions. By shifting costs of harm, the rules on liability may influence the design choices of producers in delegating decisions to AI-based systems or humans. This should be considered when choosing between a fault-based or strict liability regime.

Compared to a fault-based regime, strict liability has the advantage that legislators and courts do not need to have information on the optimal level of precaution in designing and testing AI-based solutions. By shifting the total costs of harm on injurers, a strict liability rule incentivises injurers to reduce their activity level in cases where AI applications are inherently risky, even if proper precautions are taken. However, if AI's superior performance reduces harm to society, strict liability proves an obstacle as it can reduce the beneficial use of AI applications below the efficient level. Further, strict liability might lead to socially insufficient innovation if the innovator does not internalise all the social benefits from the creation. Nevertheless, if an individual risk is highly correlated and a failure is a rare event – a possible high-risk environment – practical considerations make strict liability the preferred option.

Finally, several parties are involved in providing the product or service for many AI-based solutions. If care by each party is essential to avoid a failure, and courts cannot verify the source of the failure, even strict liability leads to a socially inefficient level of care when no punitive damages are allowed. Users must be held liable next to producers as well. Even if they exercise less control over how an AI system operates, they decide when to employ it and benefit from its use.

5. Adapting EU liability rules for AI

Section 3 demonstrated the gaps in liability rules resulting from the characteristics of AI. Section 4 gave guiding principles for how to fill these gaps to achieve efficient liability for AI-based harm. Section 5 discusses the concrete changes this would require and evalutes the EU proposals for a new PLD and an AI-specific liability rule for operators. These EU proposals determine how responsibility for AI-related harm is distributed between producers under the revised PLD and operators under the new AI Liability Directive.

5.1. *Liability for AI producers under the revised PLD*

The liability of AI producers will be determined by the revised PLD. The PLD has a horizontal scope, and the rationale for updating it is broader than the concerns identified in relation to AI. Still, this rationale is closely related to technological development. The scope of the PLD not only determines the allocation of responsibility between producers and operators but, of course, also sets the standard of care for producers.

Three issues are of main concern with regard to product liability for AI systems: the first is the scope of the concept of "product", particularly whether it includes intangible items and where to draw the line between products and services; the second is the definition of "defect" and how to interpret this for autonomous AI systems; and the third is the burden of proof. Besides those three main issues analysed below, it is interesting to note two important welcome expansions proposed by the Commission for the revised PLD. First, the Commission proposes to replace the term "producer" with "manufacturer" to include providers of software, providers of digital services and online marketplaces as possible liable parties under the PLD. As explained above in Section 4.3, the extension of the parties being liable is efficient. Second, the Commission proposes to expand the categories of harm which may be compensated to include the loss or corruption of data, such as content deleted from a hard drive.[100] It also proposes to include medically recognised damage to psychological health as part of personal injury damage.[101]

5.1.1. *Definition of product*

Currently, the PLD covers tangible products. This means that hardware components of an AI system and software integrated into tangible AI systems are covered by the PLD,[102] but

[96] The Expert Group Report (n 29) defines as operator as "the person who is in control of the risk connected with the operation of emerging digital technologies and who benefits from their operation". The proposed AI Act defines operator as "provider, the user, the authorised representative, the importer and the distributor" (COM (2021) 206 final, art 3 (8)).

[97] See also Galasso and Luo (n 98).

[98] Galasso and Luo (n 98).

[99] Benhamou and Ferland (n 51).

[100] Article 4, under (6) and Recital 16 PLD Proposal.

[101] Recital 17 PLD Proposal. The Commission did consider extending the categories of compensable damage to damage resulting from fundamental rights infringements, such as data protection breaches, privacy infringements or discrimination, but decided not to pursue this option: Explanatory Memorandum PLD Proposal, p. 9.

[102] Jessica S. Allain, 'From Jeopardy! to jaundice: the medical liability implications of Dr. Watson and other artificial intelligence systems' (2013) 73 Louisiana Law Review 7; Susanna Navas, 'Robot Machines and Civil Liability' in Martin Ebers and Susanna Navas (eds), *Algorithms and Law* (Cambridge University Press 2020), 167 w.r.t. robots; also referring to Duncan Fairgrieve, Geraint Howells, Peter Møgelvang-Hansen, Gert Straetmans, Dimitri Verhoeven, Piotr Machnikowski, André Janssen, and Reiner Schulze, 'Product liability directive' in Piotr Machnikowski (ed), *European Product Liability: An analysis of the state of the art in the era of new technologies* (Intersentia 2016) 47. Case law and jurisprudence has largely already taken this approach. Given that the Directive covers electricity, it could be argued that a product does need to be tangible, see Ebers (n 51), 58.

for standalone software, this is unclear.[103] For software, the medium can be decisive for qualifying as a product. As a result, software stored on a DVD or a Flash-drive[104] is covered by the PLD, but downloaded software may not be. Member States have applied the concept of software from the PLD differently in their national implementations.[105] Clarifying the definition of product is necessary because it determines if manufacturers of software can be held liable under the PLD.[106]

In the age of digitalisation, differentiations between tangible and intangible objects of use are more difficult to justify.[107] It is unclear why the mode in which computer programs are stored, copied, and distributed should be relevant for the application of the PLD. The main purpose of the PLD was to ensure a fair distribution of the risks associated with industrially manufactured between the injured party and the manufacturer.[108] The risks associated with downloaded software do not appear very different from their traditional counterparts supplied on CDs.[109] Once the software is introduced to a computer, it brings about material and tangible changes.[110] This is obvious where software is integrated into a machine[111] but is also easily imaginable for intangible software: one could think of an insulin therapy app used by a patient making an error or malware corrupting all of a consumer's files. The risks involved in software, irrespective of its medium, therefore support including software in the notion of products.[112] The PLD proposal recognises this and explicitly includes intangible software and digital manufacturing files as products.[113] The question remains how to delineate products from services. Under the proposed PLD, services related to products, such as a digital service interconnected with a product that is required for the product's functions to be performed, would be considered to be part of a product.

However, consumers increasingly consume items fully as a service that they used to purchase as a product.[114] For instance, where consumers would previously buy a CD, they now have a subscription to a service such as Spotify. Digital goods have blurred the distinction between products and services.[115] As cloud-computing abilities improve, more AI systems may be operated on service models as well – not just digital goods but physical ones as well. As a result, it may become increasingly difficult to draw a sharp line between products and services for digital goods and AI systems. Traditionally, the different risks associated with products and services justify imposing strict liability for product defects but not for services. A consumer of a service may more easily prove that the service provider was negligent than a consumer can provide evidence about the defective nature of a product.[116] The distinction between products and services may be less justified with respect to many digital goods, given that their risks may well be the same.[117] In the long term, either a common liability regime may have to be adopted for both or, at least, clear definitional criteria will need to be developed.[118]

5.1.2. *Definition of defect*

The definition of "defect" is pivotal in determining producer liability for autonomously operating systems. Under the current PLD, producers are only liable for a defect arising before the product was placed on the market. This reflects that producers have no control over the product from that moment onward. This control criterion is problematic if product safety relies on a producer's updates to the software[119] or if AI sys-

---

[103] See further Lutter (n 97) 282. Some authors take the position that the PLD already now extends to digital content, e.g. Bernhard A. Koch, 'Product Liability 2.0 – Mere Update or New version?' in Sebastian Lohsse, Reiner Schulze, and Dirk Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Nomos 2019) 106; Gerhard Wagner, 'Produkthaftung für autonome Systeme' (2017) 217 Archiv für die civilistische Praxis 707, 717-8; and Gerald Spindler, 'Haftung im IT-Bereich' in Lorenz Egon (ed), *Karlsruher Forum 2010: Haftung und Versicherung im IT-Bereich* (Verlag Versicherungswirtschaft 2011) 41-43

[104] Written Question No 706/88 by Mr Gijs de Vries to the Commission: Product liability for computer programs, Official Journal [1989] OJ C 114/42

[105] See e.g. Kristin Nemeth and Jorge Morais Carvalho, 'Time for a Change? Product Liability in the Digital Era' (2019) 8 Journal of European Consumer and Market Law 160 on the differences between the German and the Austrian implementation.

[106] See further Michael Stöber, Marc-Christian Pieronczyk, and Annelie Möller, 'Die Schadensersatzhaftung für automatisierte und autonome Fahrzeuge' (2020) 90 Deutsches Autorecht 609, 612.

[107] See also Stöber et al (n 109) 613.

[108] Recital paras. 2 and 7 PLD.

[109] Wagner, 'Robot, inc.: Personhood for autonomous systems?' (n 96) 604.

[110] K. Alheit,'The applicability of the EU product liability directive to software' (2001) 34 Comparative and International Law Journal of Southern Africa 188.

[111] Alheit (n 113), 201.

[112] The view that the PLD should apply independent of the mode in which computer programs are stored, copied and distributed was shared by Expert Group Report (n 29). See also Christoph Schmon, 'Product Liability of Emerging Digital Technologies' (2018) 3 Zeitschrift für Internationales Wirtschaftsrecht 254; Stöber, Pieronczyk, and Möller (n 109) 613; Hans Steege, 'Auswirkungen von künstlicher Intelligenz auf die Produzentenhaftung in Verkehr und Mobilität' (2021) Neue Zeitschrift für Verkehrsrecht 6, 7; Wagner, 'Robot Liability' (n 82) 42; Daily Wuyts, 'The product liability directive – more than two decades of defective products in Europe' (2014) 5 Journal of European Tort Law 1, 6; Rolf H. Weber, 'Liability in the Internet of Things' (2017) 3 Journal of European Consumer and Market Law 207, 210.

[113] Article 4 under (1) and Recital 12, Preamble. Recital (13), however, excludes non-commercially used, free and open-source software from the PLD's scope.

[114] Omri Rachum-Twaig, 'Whose Robot Is It Anyway?: Liability for Artificial-Intelligence-Based Robots' (2020) University of Illinois Law Review 1141, 1157 and the text in footnote 93.

[115] Expert Group Report (n 29), 28.

[116] Ebers (n 51), 58. See also Case C-495/10 *Dutreux* [2011] ECR I-14155; Commission Evaluation COM (2018) 246 final 7. See e.g. Brigid M. Carpenter and Caldwell G. Collins, 'The Shirt Off My Back: Using the Relationship Between a Product and a Service to Your Advantage' (*IADC Committee November 2021 Newsletter,* November 2012) <https://www.bakerdonelson.com/files/Uploads/Documents/ProductLiabilityNovember2012.pdf> accessed 6 December 2022.

[117] Benhamou and Ferland (n 51) 13.

[118] Scott Marcus, 'Liability: When Things Go Wrong in an Increasingly Interconnected and Autonomous World: A European View' (2018) 1 IEEE Internet of Things Magazine 4.

[119] Expert Group Report (n 29), 28; BEUC, 'AI RIGHTS FOR CONSUMERS' (BEUC-X-2019-063, 23 October 2019) <www.beuc.eu/publications/beuc-x-2019-063_ai_rights_for_consumers.pdf> accessed 07 August 2022

tems are intended to continue learning once they are placed on the market.[120] Therefore, one issue in the PLD reform was to consider the dynamic nature of software products, digital goods and AI systems.[121]

Under the proposed PLD, the test for determining whether a product is defective remains substantively the same. The proposal does allow for additional factors to be considered by courts when assessing defectiveness, such as the interconnectedness or self-learning functions of products and a product's cybersecurity vulnerabilities. The concept of defect is also broadened in that liability is no longer assessed only by reference to when a product was put into circulation. If a manufacturer retains control of a product, for instance, through software updates or machine-learning algorithms, the time after a product has been placed on the market can be considered for liability under the PLD.[122]

An obligation to monitor a product after it has been placed on the market is alien to the current PLD. The extension to defect in the PLD proposal incorporates suggestions from commentators to clarify the monitoring duties of producers for AI systems with learning capabilities.[123] It extends liability to producers that fail to provide updates relevant to the safety of the product. It moreover clarifies, as was also suggested by commentators,[124] that the failure of a user to install an update precludes liability of the manufacturer.[125] To prevent an overbroad and open-ended liability, it may be necessary to include clear criteria on how long such an obligation should reasonably exist.[126] Also, it may be helpful to clarify whether consumers may expect updates to be delivered throughout the product's life cycle.[127]

The PLD does not provide much guidance on applying the concept of defect to autonomous AI systems. If an AI system is supposed to work autonomously, the question arises if any instance of harm constitutes a defect or if it is accepted that a well-functioning AI system could still cause damage. If *some* failure is accepted, the question is *which* failure is acceptable. It is unclear how far the concept of defect extends for deliberate but undesirable operations of AI systems with self-learning capacities. It may not be possible to draw the line between harm resulting from AI's autonomous decisions and harm resulting from a defect.[128]

One option would be to extend the concept of "defect" for fully autonomous AI applications to any harm they cause. Another is to distinguish more clearly between different types of defects. This could be justified by the fact that the producer who designs the learning process for the AI system is best placed to judge whether the product is safe enough to be put on the market, and profits from selling it.[129] Such a rule would encourage producers to inform users about contexts in which the application cannot work entirely autonomously. However, from a practical perspective, if products with a higher autonomy level are treated differently under product liability, this will likely affect how products are marketed or designed. Under such a liability rule, we can expect producers to avoid liability by simply not marketing applications as fully autonomous. Producers would likely add extensive product manuals outlining the contexts in which users still have a duty to monitor the system.

Aside from this practical problem, it would be unreasonable to require absolute safety in the context of liability.[130] Certain situations, such as in healthcare, may require absolute security because of the high stakes involved,[131] which is reflected by regulatory safety standards. Generally, extending strict liability to AI manufacturers, so they are responsible for any AI harm, shifts an undue portion of the burden on manufacturers.[132] Such a regime would force AI manufacturers to bear the negative externalities without compensation for the value of the positive externalities of AI which may be substantial.[133] As explained in Section 4.3 above, it would also place an insufficient burden on the owners who benefit from employing AI and impose risks on others by doing so (see further below). Moreover, waiting for nearly perfect AI before using it is likely more costly than accepting a reasonable failure rate. The liability rules should reflect this.[134]

A second option would be to differentiate producer liability depending on the type of defect, similar to US product liability law.[135] For instance, strict liability could be limited to manufacturing defects, while a presumption of fault could be applied for defects in design and instructions to users.[136] A consequence of this approach is to limit liability for AI systems compared to other products if AI systems are less vulnerable to manufacturing defects and more susceptible to design flaws.[137]

[120] Commission AI Report, p. 15. See also Seehafer and Kohler (n 97) 214.
[121] Expert Group Report (n 29) 43.
[122] Article 4(5) PLD proposal.
[123] Seehafer and Kohler (n 97) 214, referring to Tim Hey, *Die außervertragliche Haftung des Herstellers autonomer Fahrzeuge bei Unfällen im Straßenverkehr* (Springer Gabler 2019).
[124] Seehafer and Kohler (n 97) 217; Steege (n 115) 12. The Commission already noted that subsequent updates cannot be the sole responsibility of the manufacturer: the user would have the obligation to install safety-relevant updates (COM (2020) 64 15).
[125] Recital 38 PLD proposal.
[126] Steege (n 115) 12.
[127] Schmon (n 115); Seehafer and Kohler (n 97).
[128] Benhamou and Ferland (n 51) 7.
[129] See also Caroline Cauffman, 'Robo-liability: The European Union in search of the best way to deal with liability for damage caused by artificial intelligence' (2018) 25 Maastricht Journal of European and Comparative Law 527, 530.
[130] See e.g. PT Schrader, 'Haftungsfragen für Schäden beim Einsatz automatisierter Fahrzeuge im Straßenverkehr' (2016) 86 Deutsches Autorecht 242, 243; and Friedrich Graf von Westphalen, 'Produkthaftungsrechtliche Erwägungen beim Versagen Künstlicher Intelligenz (KI) unter Beachtung der Mitteilung der Kommission COM(2020) 64 final' (2020) 35 Verbraucher und Recht 248, 250.
[131] Lutter (n 97) 283.
[132] Yoshikawa (n 70) 1165 and 1171.
[133] See Yoshikawa (n 70).
[134] Nidhi Kalra and David G. Groves, *The Enemy of Good: Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles* (RAND Corporation 2017) <www.rand.org/pubs/research_reports/RR2150.html> accessed on 16 December 2022.
[135] See further Wuyts (n 115) 10.
[136] Navas (n 105) 168.
[137] Navas (n 105); F. Patrick Hubbard, 'Sophisticated robots: balancing liability, regulation, and innovation' (2014) 66 Florida Law Review 1803, 1821-1823; Martin Ebers, Autonomes Fahren:

A third possibility is to treat design defects in the same way as a manufacturing defect and impose a strict liability regime.[138] As it may be difficult to identify a design error by looking at the individual AI system, one could consider the harm caused by a fleet of AI systems that operates by the same algorithm to determine whether there is a defect in design. It then still needs to be determined what failure rate is deemed acceptable – generally, the safety requirements placed on the manufacturer increase with the risks associated with the product. We may also expect AI systems to be safer than the "dumb" products they are replacing. However, we need to consider how much safer than human decision-making we require AI systems to be. Using the safest technology in the market as a benchmark would essentially banish the AI systems offered by all but one competitor from the market. The PLD Proposal recognises this problem, stating that a product is not considered defective "for the sole reason that a better product [...] is placed on the market."[139]

At the same time, it is not helpful to compare the performance of an AI system with how a carefully acting human would have behaved in a specific situation. The first reason is that a comparison with good human decision-making is pointless because we want AI to do better than humans.[140] Second, the point of reference differs: in the case of a human being, the reference point is the decision to act in an individual case, while for an AI system, it is whether the programming for an entire series of products could and should have been done more carefully to prevent the occurrence of the damage.[141] Courts need to identify shortcomings that could have been avoided by alternative programming.[142] Self-learning AI systems that initially function well and develop a malfunction in practical use could be considered already initially not error-free.[143] A third reason is that the pool of accidents that an autonomous system causes might be easily avoidable by humans – one can think of the ability of an autonomous car to recognise a white truck in a bright environment. Despite making errors that humans would not, AI systems may overall still make significantly fewer errors. As a result, it may be misguided to compare the standard for safety to humans.[144]

Overall, with regard to autonomous AI systems, we need to consider what design flaws for AI are unacceptable and what error rate is unacceptable. Moreover, the burden of proof (discussed below) and regulatory safety standards may help mitigate the challenges that autonomous AI systems pose for the concept of the defect.

Regulatory safety standards can also help reduce unintended negative consequences from autonomous decision-making. It is essential to increase our understanding of how AI systems learn once they are placed on the market to take the appropriate regulatory steps. A product defect becomes more difficult to recognise or even define if AI devices continue to learn independently once they are on the market. Such devices would be less predictable and harder to control, even for manufacturers. If AI devices are thoroughly tested after a learning process and "frozen" when placed in the market, harm from unintended actions may be less likely. If regulation precludes AI products from entering the market without "freezing" them, this could reduce the need for interpreting the concept of defect more broadly in the product liability rules. However, such an option should always be weighed against the lost benefits of not employing these systems with learning capabilities.

5.1.3. *Burden of proof*

The PLD requires injured parties to prove that the product was defective and that it caused the injury. Proving a defect can be difficult for consumers for any technically complex product. National courts have therefore developed ways to facilitate the burden of proof in such situations, including by disclosing obligations for the producer or allocating the costs of experts' opinions.[145]

The PLD Proposal does not reverse the burden of proof, as this was considered to expose manufacturers to too high liability risks. Nevertheless, the proposal does alleviate the burden of proof in some circumstances. As discussed in Section 3, the proposal introduces a rebuttable presumption of a causal link between the defendant's fault and the output (or lack thereof) produced by the AI system.

This can facilitate claimants in product liability cases involving AI systems. Proving a defect may be more complex for AI systems, as the defect may be difficult to identify. If, for instance, an AI diagnosis tool delivers a wrong diagnosis, there may be no apparent malfunctioning to the user.[146] Depending on the definition of a defect, users may be asked to show that harm resulted from a flaw in the AI device and not from its autonomous decision-making. Proving causality in the context of AI harm may be difficult, especially if some human supervision is still required.[147] The injured party may have difficulty proving that the AI system, not their own negligence, caused the harm. AI developers may also argue that it is impossible to precisely anticipate how AI systems will act, meaning that the harm is unforeseeable.[148] While this is unlikely to succeed as a defence, such questions could arise when AI did exactly what it was intended to do (act autonomously) and nevertheless caused harm. The assessment of the causal link will of-

Produkt- und Produzentenhaftung in Bernd H. Oppermann and Jutta Stender-Vorwachs (eds), *Autonomes Fahren* (C.H. Beck 2017) 111-12.

[138] Mark A. Lemley and Bryan Casey, 'Remedies for robots' (2019) 86 The University of Chicago Law Review 1311, 1327-8.

[139] Article 6(2) PLD proposal.

[140] Jean-Sébastien Borghetti, 'How can Artificial Intelligence be Defective?' in Sebastian Lohsse, Reiner Schulze, and Dirk Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Nomos 2019) 69.

[141] Seehafer and Kohler (n 97)) 214 and footnote 23 therein.

[142] Wagner, 'Robot, inc.: Personhood for autonomous systems?' (n 96).

[143] See Philipp Etzkorn, 'Bedeutung der «Entwicklungslücke» bei selbstlernenden Systemen – Rechtliche Fragen zur fortdauernden Softwareentwicklung durch maschinelles Lernen im Praxiseinsatz' (2020) 23 Multimedia und Recht, 360, 362.

[144] Wagner, 'Robot, inc.: Personhood for autonomous systems?'(n 96) 605.

[145] Wuyts (n 115) 24.

[146] Borghetti (n 143) 67.

[147] Causality is governed by national rules, given that the PLD does not define a causal relationship. See on different national approaches e.g. Wuyts (n 115) 25.

[148] Kowert (n 71) 191.

ten require expert advice, the cost of which may discourage injured parties from suing.[149]

While reversing the burden of proof for parties injured by highly complex technologies has been suggested in the literature,[150] this would have significantly altered the distribution of risks to the detriment of manufacturers. It would also have departed sharply from the current principles of the PLD.[151] Thus, the rebuttable presumption of causality appears to be a good compromise. Given that AI systems may be equipped with event logging or recording systems, victims may also get access to better data about the cause of an accident than non-AI products.[152] On the national level, perhaps the rebuttable presumption in the PLD will be accompanied by evidence disclosure duties,[153] cost-shifting rules for expert advice, or – insofar as data protection rules permit – requirements to collect data about the functioning of the system. This would allow manufacturers to retrace possible causes for an error at a later stage.[154]

5.2. *Liability for AI operators under the proposed AI Liability Directive*

Sections 3 and 4 demonstrated that fault-based liability of operators may not work effectively for AI systems and that it would be efficient to hold operators liable. The European Commission has proposed to introduce a stricter standard of care for operators through a rebuttable presumption of causality and to allow claimants access to relevant information. With some exceptions, these rules apply to high-risk AI as defined in the proposed AI Act.

5.2.1. *Scope of a specific AI liability regime*

The European Commission has proposed a horizontal liability framework for pre-defined high-risk AI systems which would complement the safety rules in the proposed AI Act. In practice, it will need to be sufficiently clear for users (or other types of operators) and courts to understand what applications are covered by this framework. Civil cases will likely revolve around the question of whether a specific device is AI and whether it is a high risk as this will determine whether it is covered by the general fault-based national liability regime or by the stricter liability rules of the proposed AI Liability Directive.[155] To avoid introducing a new source of legal uncertainty, the proposed AI Liability Directive, together with the proposed AI Act, needs to clearly define which AI applications are high-risk. The ongoing debate to define high-risk AI in the proposed AI Act[156] illustrates that this is not straightforward. The definition in the proposed AI Act focuses on the types of harm involved (e.g. health and fundamental rights) and the areas in which AI systems are employed. As Section 4 illustrates, from an efficiency standpoint, other aspects of risks should be taken into account, such as whether risks are highly correlated.

One advantage of the horizontal option chosen by the European Commission is to be more flexible and adaptable to technologies that evolve quickly.[157] As AI applications differ in both the benefits and the risks they create for society, it is appropriate to differentiate between the regulatory and liability requirements that apply to different AI applications. However, we identify potential problems with introducing a horizontal liability framework for high-risk AI applications. Listing "high-risk" AI applications may presuppose that AI applications create similar risks regardless of the context in which they are applied. AI encompasses various technologies, which may be used in a wide range of applications, which in turn could be employed in various contexts. Courts will therefore have to review the risk not only for the particular device but, particularly for general-purpose devices, also the particular use of that device.

A narrower and more focused option possibility would have been to align the scope of a stricter EU liability standard with existing EU sector-specific rules.[158] Existing sector-specific regulation already reflects the need to differentiate regulation according to the context in which technology is applied. The advantages would be that the scope would already be legally defined – for instance, transportation or medical devices – and that a coherence between safety and liability rules could be achieved. From a legitimacy point of view, this option also ensures that the scope of the liability rule is defined by the legislator when adopting sector regulation and not by the courts when interpreting criteria to define high-risk AI applications.

In practice, it remains to be seen how many AI applications will be covered by the liability regime of the proposed AI Lia-

[149] Cauffman (n 132) 530; Wuyts (n 115) 24.

[150] Charlotte De Meeus, 'The Product Liability Directive at the Age of the Digital Industrial Revolution: Fit for Innovation?' (2019) 8 Journal of European Consumer and Market Law 149, 151; Wagner, 'Produkthaftung für autonome Systeme' (n 106) 747; Seehafer and Kohler (n 97) 216; Roeland de Bruin, 'Autonomous Intelligent Cars on the European Intersection of Liability and Privacy' (2016) 7 European Journal of Risk Regulation, 485, 495, who acknowledges that this could negatively impact innovation.

[151] See further Seehafer and Kohler (n 97) 216.

[152] Koch (n 106) 110; Gerald Spindler, 'Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien?' (2015) 31 Computer und Recht 766, 772.

[153] Schmon (n 115) 6.

[154] Antje von Ungern-Sternberg, 'Artificial Agents and General Principles of Law' in Andreas von Arnauld, Kerstin von der Decken, & Nele Matz-Lück (eds), *German Yearbook of International Law* (Duncker & Humblot 2018) 9.

[155] See also Sebastian Lohsse, Reiner Schulze, and Dirk Staudenmayer, 'Liability for Artificial Intelligence' in Sebastian Lohsse, Reiner Schulze, and Dirk Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Nomos 2019) 21

[156] See e.g. Luca Bertuzzi, 'Leading MEPs exclude general-purpose AI from high-risk categories – for now', Euractiv, 12 December 2022, < https://www.euractiv.com/section/artificial-intelligence/news/leading-meps-exclude-general-purpose-ai-from-high-risk-categories-for-now/ > accessed on 16 December 2022; Khari Johnson, 'The Fight to Define When AI Is 'High Risk'', Wired, 1 September 2021, < https://www.wired.com/story/fight-to-define-when-ai-is-high-risk/> accessed on 16 December 2022.

[157] It was also favoured by the European Parliament in its Resolution of October 2020 which recommends that all high-risk AI-systems be exhaustively listed in an Annex to Regulation it proposes. See EP Resolution, art 4.

[158] Buiten (n 48); Chris Reed, 'How should we regulate artificial intelligence?' (2018) Philosophical Transactions of the Royal Society A 376:2128.

bility Directive. On the one end, there is a large group of AI systems that is not high-risk. For this group, a stricter liability regime is not justified if we broadly follow the existing principles for liability in the Member States. The proposed AI Liability Directive rightfully largely limits its scope to high-risk AI systems. On the other end, some applications are already covered by specific liability rules, such as autonomous vehicles. These AI applications can continue to be governed by sector-specific rules. The proposed AI Liability Directive indeed excludes liability in the field of transport from its scope.[159] Some Member States may find it useful to extend the liability regime for transport to certain AI systems, for instance, drones. For such devices, compulsory liability insurance schemes may also need to be imposed, similarly to car owners.[160] If such rules are introduced for many categories of high-risk AI systems or in several sectors, the set of AI systems that is high-risk but is not regulated may end up being small.

The question is also to what extent the characteristics of AI justify more EU harmonisation of liability rules beyond the context of the PLD. The level of harmonisation of liability rules and the scope of such harmonised rules present trade-offs. On the one hand, harmonisation may help ensuring legal certainty for injured parties and operators with a uniform framework. On the other hand, not harmonising liability rules may help preserving the internal coherence of Member States' national liability rules and allow for learning effects in the Member States.

5.2.2. *AI liability standard*

The European Commission choses a limited approach in the proposed AI Liability Directive, opting for a rebuttable presumption of causality rather than introducing a higher duty of care or strict liability for operators of high-risk AI systems. From the perspective of victims, this may be the preferred solution – potentially even to strict liability, because it solves the need to prove causality. As explained in Sections 3 and 4 above, the presumption may help to establish causality where AI systems are opaque or autonomous, and the operator has little control.[161]

A procedural mechanism is also easier to define at the EU policy level than a standard of care for operators that is sufficiently clear to be uniformly interpreted by national courts while allowing sufficient flexibility to deal with a wide variety of AI systems and contexts. Indeed, under the current proposal, the rebuttable presumption would be applied to the varying liability standards in the Member States. Member States that have broad strict liability rules in place, such as France, would continue to apply strict liability in contexts covered by the proposed AI Liability Directive. It remains to be seen what this will mean for the application of the rebuttable presumption of causality in practice and what level of uniformity for liability throughout the EU can really be achieved.

Moreover, under the proposed approach, national courts would still need to specify a duty of care of the operator.[162] The standard of care for operators could consist of the operator's choice to employ an AI system which is reasonably safe, maintain it, and monitor or supervise it.[163] In turn, the operator's duties to monitor an AI system would depend on the level of autonomy of the AI system as well as its level of risk. If the operator may reasonably expect the AI system to act fully autonomously, the operator has no duty to monitor an AI system.[164] The duty of the operator is then limited to how she maintains and employs the AI system. We can expect many, if not most, AI systems to require some level of oversight from humans still. In cases of harm, courts would then need to specify if the operator was at fault for failing to intervene or override the AI system.

In setting a standard of care for operators, the instructions of producers on using and supervising an AI system can act as guidance to courts. Producers have an interest in providing precise warnings and instructions to users because cases of harm might open them up to producer liability as explained above. This is true for any product but particularly relevant for AI systems for which users need to know if and how they need to monitor it. Producer liability thus promotes information disclosure by producers. This information could serve as a benchmark for the standard of care for the operator. For instance, manuals for vacuum robots warn against employing the device in the presence of small children.[165] If an operator of an AI system fails to follow these instructions and harm occurs, this could be an indication that the operator was at fault and should be held liable.

There are some limits to the potential for information disclosure, however. Long lists of warnings are likely to be ignored by consumers in the same way that general terms and conditions are not read.[166] To ensure that consumers are effectively informed about AI systems, standards on information disclosure may need to be set. It may be useful to introduce "autonomy labels" for AI, akin to the European energy labels. The autonomy labels could be aligned with certification processes and other safety regulations and would indicate to consumers what level of supervision is required when using an

---

[159] Article 1(3)(a) AI Liability Directive Proposal.

[160] See e.g. Georg Borges, 'Product Liability 2.0 – Mere Update or New version?' in Sebastian Lohsse, Reiner Schulze, & Dirk Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Nomos 2019); Navas (n 105) 166; David Levy, 'Intelligent no-fault insurance for robots' (2020) 1 Journal of Future Robot Life 35.

[161] See also A. Lior, 'AI entities as AI agents: Artificial intelligence liability and the AI Respondeat Superior Analogy' (2020) Mitchell Hamline Law Review, 46, 1044.

[162] For a proposal (in the U.S. context) see Rachum-Twaig (n 117) 1168-70.

[163] Expert Group Report (n 29), 44; Ruth Janal, 'Extra-Contractual Liability for Wrongs Committed by Autonomous Systems' in Martin Ebers and Susana Navas (eds), *Algorithms and Law* (Cambridge University Press 2020) 193 notes that, "the users of an autonomous system may be held liable for the acts of the system if they have breached a duty of care, particularly in operating and supervising the autonomous system."

[164] See also Janal (n 166) 193.

[165] See <https://prod-help-content.care.irobotapi.com/files/s_Series/s9/ownersGuide/ownersGuide_enUS.pdf> accessed on 5 August 2022.

[166] See CERRE Report by Alexandre De Streel and Anne-Lise Sibony, Towards Smarter Consumer Protection Rules for the Digital Society (*CERRE Policy Report*, 5 October 2017) <https://cerre.eu/wp-content/uploads/2020/06/171005_CERRE_DigitalConsumerProtection_FinalReport.pdf> accessed on 07 August 2022.

AI application. Given that the autonomy labels would provide information about the delegation of decisions to the AI system and the humans involved, these labels can inform courts when assigning liability to producers, operators and users. As already mentioned above, a drawback of this could be that producers may be discouraged from developing AI systems with increased autonomy if this increases their liability, even if this would be a safer alternative to semi-autonomous systems that still require human oversight in crucial situations. Nevertheless, autonomy labels could help resolve information problems of courts by setting clear standards for the division of responsibility for harm involving AI systems.

Overall, as AI systems reach higher levels of autonomy, the duty of care of the operator may become more difficult to establish in concrete cases. At some point, AI systems are arguably no longer tools used by humans but rather machines deployed by humans that act independently of direct human instruction.[167] This may require legal rules by analogy to the liability of employers for employees or of parents for their children. It could involve imposing strict liability on operators of a selected group of AI systems, for instance, in sector regulation.

A strict liability standard may have several advantages as shown in Section 4.2 above. First, the advantage of strict liability to establishing a "duty to supervise" under fault-based liability is that it ensures compensation for victims also in cases where, even if operators monitor an AI system, they may not be able to prevent harm if an AI system acts in completely unexpected ways.[168] AI systems with particularly high risks, it could be justified to decouple liability from fault rather than raising the standard of care or reversing the burden of proof. Second, strict liability can save the high transaction costs that injured parties would need to incur to litigate liability issues involving autonomous systems where the fault is difficult to establish.[169] Third, a strict liability regime may be more predictable. It would likely lead to fewer interpretation variations across national courts in the Member States.

Interestingly, a strict liability rule for (some) AI systems would have analogies in the existing liability regimes of several Member States. Following the typology of the human-AI relationship, AI liability could follow the liability of parents for their children, owners for their animals or principals for their agents. Under a type of parental liability, operators would evade responsibility only if they could prove it was not possible to prevent a machine's action.[170] A type of employer liability could be justified since, by employing AI systems, operators impose risks on others. Particularly where a corporation operates an AI system, we may think of the corporation as operating the robot on its behalf.[171] By holding operators strictly liable for their AI systems as a principal, they are properly incentivised to take precautions and identify the optimal level of employing the AI system.[172] The third possible analogy is the owner-animal relationship: In their erratic and unpredictable behaviour, AI systems resemble animals.[173]

Liability for animals or children reflects the risk emanating from their independent, not fully controllable behaviour. Strict liability is usually justified with the consideration that a particular danger arises from certain useful and therefore permitted facilities or activities. Those persons who are served by the facility or activity should also be assigned the disadvantages caused.[174] While the actions of advanced AI systems may similarly be uncontrollable, they usually promise a significant increase in safety as compared to their non-AI or human counterparts.[175] For this reason, AI systems may be of important value to society and imposing strict liability on their operators may impose an excessive burden.[176] In the case of the principal-agent relationship, strict liability has a different justification. The key consideration here is that parties shift risk to another party and should bear the consequences of their actions. This argument also holds in the AI context, even if defining "control" over the system may be challenging.[177]

6. Conclusion

As AI technologies enter everyday products and services, they are bound to play a role in liability lawsuits as well. This raises the question of whether liability rules are apt to deal with AI. This paper identified several possible gaps in liability rules, analysed the efficient liability regime for AI, and evaluated the recent EU proposals for producer and operator liability.

The paper identified three dimensions relevant to reviewing the EU liability framework for AI systems: (1) the scope of liability rules, (2) the liable parties; and (3) the standard of liability. The European Commission has proposed to adapt liability rules for producers on a horizontal level with the revision of the PLD, while introducing AI-specific liability rules for users and owners of these systems. The broad review of the PLD illustrates that many of the challenges often associated with AI are in fact much more general, relating to digital goods and services. One can think of the involvement of many stakeholders and data-drivenness. At the moment, the unique challenges of AI likely affect only a narrow group of products and services: those AI systems that exhibit autonomous and hence unpredictable characteristics. It remains to be seen how much the AI Liability Directive, if adopted, will be applied, if many high risk AI systems will be covered by sector specific liability rules. It is also to be seen how the rules will

167 David C. Vladeck, 'Machines without principals: liability rules and artificial intelligence' (2014) 89 Washington Law Review 117, 121.
168 Janal (n 166) 199.
169 See e,g, Gerald Spindler, 'Zukunft der Digitalisierung – Datenwirtschaft in der Unternehmenspraxis' (2017) Betrieb (DB) 41, 50.
170 Pagallo (2012) 56.
171 Rachum-Twaig (n 117) 1151.
172 Lior (n 164) 4.
173 Lior (n 164) 18. The liability standard for animals varies across the Member States, and within some Member States, e.g. depending on the purpose of the animal/AI system. See with respect to Germany e.g. Borges (2018), pp. 981-982.
174 Julian Pehm, 'Systeme der Unfallhaftung beim automatisierten Verkehr' (2018) IWRZ, 259, 265.
175 Rachum-Twaig (n 117) 1158.
176 Rachum-Twaig (n 117) 1145.
177 Rachum-Twaig (n 117) 1151. However, a challenge would be that employer liability usually requires that the principal has control over the agent, which may be more difficult to establish for unpredictable behaviour by an AI system.

play out in the different Member States, where fault-based or strict liability rules may apply to AI systems in various contexts.

As this paper has illustrated, it is useful to hold both producers and operators liable for AI, even if operators have less control over these systems. Operators still choose when to employ an AI system, can maintain and oversee it, and benefit from its use. The extent of operator liability is largely influenced by the scope of producer liability. Therefore, the reviewing of the PLD is an important step in allocating responsibility for AI systems between producers and operators. Overall, non-contractual liability should provide incentives to all stakeholders to take an efficient level of care in designing, testing and employing AI-based solutions, recognising that care by each party may be essential to avoid a failure.

Policymakers should be aware of how liability rules may affect the adoption of AI. When AI technologies are safer than their traditional counterparts, there are opportunity costs of not employing AI. From this perspective, liability rules should be technologically neutral, providing the same level of protection for users of a product or service powered by AI as users of the same type of product or service which is not powered by AI.[178] As always, with regulatory design, rules reflect trade-offs which should be well-identified. Liability rules address possible trade-offs between the interests of the producers (and their innovation) and the interests of the users (and their protection). Liability rules should be based on risks of harm, which may differ depending on the application and the context in which AI systems are used. Liability rules should also ensure an efficient disclosure of information in situations of information asymmetries between stakeholders. In terms of timing, liability rules should balance proactive policymaking, anticipating technological changes, with reactive policymaking, adapting the rules only after having gained some experience from deploying the technologies.

More fundamentally, EU non-contractual liability rules should not be thought of in isolation but as part of a broader set of rules as they jointly shape the incentives of all parties. In particular, liability rules need to be coherent with EU general regulation (such as the proposed AI Act) and sector specific safety regulations, with compulsory or self-regulated certification schemes, and with the national non-contractual and contractual liability rules and rules on insurance.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

No data was used for the research described in the article.

[178] Other reasons may lead policymakers to follow different ethics and, thus, depart from technological neutrality.