AI liability in Europe: How does it complement risk regulation and deal with the problem of human oversight?

Beatriz Botero Arcila

Computer Law & Security Review 54 (2024) 106012

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

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

AI liability in Europe: How does it complement risk regulation and deal with the problem of human oversight?

Beatriz Botero Arcila

Assistant Professor of Law, Sciences Po Law School, Faculty Associate, Berkman Klein Center for Internet and Society Harvard University, 13 Rue de l'Université, Paris, 75016, France

ARTICLE INFO

Keywords:
AI liability
AI Act
Human in the loop
Fundamental rights
AI risks
AI harms
AI bias

ABSTRACT

Who should compensate you if you get hit by a car in "autopilot" mode: the safety driver or the car manufacturer? What about if you find out you were unfairly discriminated against by an AI decision-making tool that was being supervised by an HR professional? Should the developer compensate you, the company that procured the software, or the (employer of the) HR professional that was "supervising" the system's output?

These questions do not have easy answers. In the European Union and elsewhere around the world, AI governance is turning towards risk regulation. Risk regulation alone is, however, rarely optimal. The situations above all involve the liability for harms that are caused by or with an AI system. While risk regulations like the AI Act regulate some aspects of these human and machine interactions, they do not offer those impacted by AI systems any rights and little avenues to seek redress. From a corrective justice perspective risk regulation must also be complemented by liability law because when harms do occur, harmed individuals should be compensated. From a risk-prevention perspective, risk regulation may still fall short of creating optimal incentives for all parties to take precautions.

Because risk regulation is not enough, scholars and regulators around the world have highlighted that AI regulations should be complemented by liability rules to address AI harms when they occur. Using a law and economics framework this Article examines how the recently proposed AI liability regime in the EU – a revision of the Product Liability Directive, and an AI Liability effectively complement the AI Act and how they address the particularities of AI-human interactions.

1. Introduction

In the European Union and elsewhere around the world, AI governance is turning towards risk regulation.[1] Risk regulation is a particular approach for controlling activities that create risks of harm, which relies on instruments such as standards, prohibitions, and risk and impact assessments to regulate behavior ex-ante; that is, before or at least independently of whether the potential harm actually occurs.[2] In the EU the recently approved AI Act creates a hierarchy of different levels of riskiness for AI systems, and requires the providers of high-risk AI systems to produce documentation on the functioning of the systems they deploy, comply with certain safety requirements, and participate in the creation of substantive optional safety standards.[3]

Risk regulation is a regulatory mechanism often employed when the harms potentially caused by the activities at issue are hard to disincentivize via other main instruments to control harmful activities, such as the market or liability law.[4] Even though regulation is expensive (both in terms of compliance and enforcement), economic theory justifies it when market failures allow an actor conducting a dangerous activity (such as developing and deploying high-risk AI models) to take precautions to not unduly expose society to harm. These market failures may be incomplete information by victims, consumer misperceptions

E-mail address: beatriz.boteroarcila@sciencespo.fr.

[1] See Margot Kaminksi, "The Developing Law of AI Regulation: A Turn to Risk Regulation" (*Lawfare*, April 21, 2022; https://www.lawfaremedia.org/article/the-developing-law-of-ai-regulation-a-turn-to-risk-regulation.

[2] See i.e. Steven Shavell, "Liability for Harm versus Regulation of Safety" (2004) NBER Working Paper No. 21218.

[3] European Parliament, Artificial Intelligence Act Corrigendum (19 April 2024) (AI Act).

[4] See Margot E. Kaminski, "Regulating the Risks of AI" (2023) 103 Boston University Law Review 18.

https://doi.org/10.1016/j.clsr.2024.106012

Available online 29 June 2024

about the product, or externalities that allow risk-takers to conduct their activities at a cost that is lower than their societal cost.[5] Risk regulation alone is, however, rarely optimal. The tools of risk regulation do not offer those impacted by AI systems – either in their fundamental rights or other legally protected interests - any rights and little avenues to seek redress – alone.[6] From a corrective justice perspective risk regulation must also be complemented by liability law because when harms do occur, harmed individuals should be compensated.[7] Consequently, scholars and regulators around the world have highlighted that AI regulations should be complemented by liability rules to address AI harms when they occur.[8]

The main question addressed in this Article is, thus, what should the liability rules be that complement AI risk regulation. To address it, it studies the EU's 2022 AI liability proposals, the AI Liability Directive (AILD) and a revision of the Product Liability Directive (PLD which seeks to complement the Artificial Intelligent Act (AI Act) risk and safety regulation. These proposals are an important complement to the AI Act's risk and safety approach. Indeed, relying solely on risk regulation has distributive consequences, including the possibility that individual harms and costs will be dismissed if a particular measure makes sense collectively, which may especially harm minorities.[9] It may also lead to situations where, because regulators are fallible, regulations set suboptimal standards and organizations won't have enough incentives to take optimal care.[10] Similarly, one of the main arguments that were raised when the AI Act was first published was that it didn't include individual rights nor rights of action for affected persons, even if its stated goal is to protect fundamental rights in Europe.[11] In this context, liability law becomes an important vehicle to ensure that the vast and fast adoption of AI systems in all facets of life and society is done in a way that guarantees the protection of people's rights and interests, but also to provide legal certainty for AI developers and deployers.

Using a law and economics framework this Article evaluates how the proposed AI liability regime complements the AI Act in reducing socially wasteful AI accidents by incentivizing precautionary measures, and in offering victims of harm improved avenues to seek compensation. It does so especially considering the complexity of AI and the involvement of humans in AI accidents. It finds, in a nutshell, that the AILD and the PLD, in their current forms, fall somewhat short of their ambition to effectively complement the AI Act, especially because they very strongly rely on the tiered framework developed by the AI Act. Indeed, both the AILD and PLD tend to focus on making it easier for plaintiffs of in accidents involving high-risk systems (as defined by the AI Act) to access relevant evidence or creates rebuttable presumptions that should make their burden easier. Additionally, the AILD does not apply to accidents where a human is involved in supervising the AI system. The paradox, however, is that by doing so the AILD and PLD fail to effectively complement the AI Act in the cases where it may be most useful: for systems where little or no other regulatory requirements are in place, or in some of the cases involving high-risk systems which must, according to the AI Act, be designed to be effectively supervised by a human.

The second main finding of this Article is that central issues for future liability cases such as whether a human supervisor was "effectively empowered" to supervise an AI system, and or exercising due care, will importantly depend on the standards that are yet to be developed, following the approval of the AI Act.

The ambition of this article is for these conclusions to contribute to the debate on AI liability in Europe, as well as to the broader discussion on the complementarity of risk regulation and liability law in AI governance across different jurisdictions.

This Article proceeds as follows: Part 2, is the background section. It surveys the literature on the challenges of regulating AI, the policy conversation in Europe, and the law and economics framework on the institutional choices to control the risk of harm and the complementarity of regulation and liability to address the risks of AI. Part 3 presents the two proposed AI liability directives as they relate to the framework set in place by the AI Act. Part 4 analyzes the complementarity of the AI liability directives with the AI Act, paying special attention to how, together, they facilitate victims' access to corrective justice incentivize precautionary measures, and reducing socially wasteful AI accidents, especially considering the complexity of AI and the involvement of multiple actors in AI accidents. Part 5 finishes by offering some suggestions for reform and the wider AI governance conversation.

2. Background: AI and the institutional choices to control risks of harm

This first Part outlines the now well-known specific risks posed by AI to legally protected interests and why these features complicate AI accountability when harm occurs. It then presents the theory drawing from law and economic analysis to assess the desirability for risk regulation and liability and their complementarity. Lastly, it lays out a framework to assess the complementarity of these two regimes, which will be later applied to the EU framework.

2.1. Controlling AI harms and risks: the technical and organizational challenges

There is a vast literature on the benefits and risks of AI systems.[12] It is well recognized that AI systems can enhance efficiency and productivity, and enable more accurate data analysis, aiding in better decision-making in a variety of fields.[13] At the same time, it is also well documented that AI systems pose several risks and can cause a variety of harms: AI systems like automated vehicles or appliances can pose safety risks, to life, bodily integrity, or property; AI-powered decision-making software poses risks to fundamental rights, privacy, human dignity, and equality; and AI systems also pose epistemic risks. They may, for example, slowly change how we conceptualize the world as organizations increasingly rely on profiling or sorting algorithms to make

[5] Eric Marsden, "Risk regulation, liability and insurance: Literature review of their influence on safety management," [2014] Les Cahiers de la sécurité industrielle FonCSI no. 2014-08.

[6] See EDRi et al., "An EU Artificial Intelligence Act for Fundamental Rights: A Civil Society Statement" 30 November 2021 <https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf> accessed 30 October 2023; Lilian Edwards, "Regulating AI in Europe: four problems and four solutions" (2022) Ada Lovelace Institute; Marco Almada and Nicolas Petit, "The EU AI Act: Between product Safety and Fundamental Rights" (December 20, 2022) <https://ssrn.com/abstract=4308072> accessed 30 October 2023, Kaminski (n4).

[7] Marsden (n5) 20 see also Nicomachean Ethics, Book V.

[8] See European Commission, White Paper on Artificial intelligence. A European approach to excellence and trust, Brussels, Feb. 19, 2020, COM(2020) 65 final, (2020) (White Paper on AI).

[9] Kaminski (n4), 8.

[10] Shavell (n2).

[11] EDRi and others (n6).

[12] AI is used in this piece following the definition adopted by the European Commission: "a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate output such as predictions, recommendations, or decisions influencing physical or virtual environments. Lucas Bertuzzi, "EU lawmakers set to settle on OECD definition for AI" (Euractiv, Mar. 7 2023) <https://www.euractiv.com/section/artificial-intelligence/news/eu-lawmakers-set-to-settle-on-oecd-definition-for-artificial-intelligence/> accessed 30 October 2023.

[13] White Paper on AI (n8).

decisions.[14]

What is particular about AI from a liability perspective, however, is that when harms occur AI systems' characteristics make them hard to scrutinize. AI systems have characteristics that complicate understanding, and often fully predicting, their behavior. Machine learning (ML) algorithms, for example, power many of the AI-powered tools consumers are often in contact with, such as assisted driving, healthcare, and home appliances like Amazon's Alexa,[15] and are used to make classifications, predictions and to decide what can be the best action in a particular situation.[16] ML algorithms work with high-dimension data to determine what features are relevant to that decision. The number of features can run into the tens of thousands which, even if it is replicating work done by humans, involves a qualitatively different decision-making logic from that of humans.[17] Trained machine learning algorithms define decision-making rules to handle new inputs that not need to be understood by a human operator.[18] This makes AI opaque, in the sense that recipients of the output of an algorithm rarely have a concrete sense of how the output was arrived at from the inputs – or what those inputs were.[19] They are also complex in the sense that their behavior arises in a nonlinear, often unpredictable way from that of its parts,[20] and sometimes autonomous, which comes from their mathematical optimization in high-dimensionality processes. This is also what allows their self-learning capacity.[21]

Importantly, the organizational structures in which AI systems are embedded and deployed accentuate these challenges. AI opaqueness is not only a feature of AI systems' mathematical complexity, but it can also be a function of proprietary protections of corporate or state secrecy, or because of generalized technical illiteracy.[22] Similarly, what is commonly referred to as AI socio-technical systems involve a variety of actors and elements that participate in the design of the system throughout its life cycle, program it, decide when and for what it will be adopted, and supervise it. The involvement of multiple individuals or actors in the development, deployment, and operation of AI systems is referred to as the problem of many hands and complicates assigning responsibility and accountability for AI outcomes.[23] This is especially the case if the AI provider is not the same actor as the person at issue or their employer.

This latter issue, the problem of many hands, deserves some additional discussion for its relevance for liability.

2.2. Controlling AI harms and risks: the problem of many hands

The challenges brought about by AI systems' technical and organizational characteristics are aggravated as humans interact with AI systems. Early conversations about the regulation and liability of AI focused on the "substitution effect:" What the law should do when an AI system replaces a human actor such as a driver, a decisionmaker, or a medical doctor.[24] The development and best practices around these tools today, however, reveal that AI development seems to be oriented towards situations where often, humans and AI systems collaborate.[25] In addition, regulations increasingly a *mandate* that a human is involved in different forms of AI decision-making processes; so-called human-in-the-loop.[26]

The key assumption of these human-in-the loop mandates is that humans and machines can complement each other well: Algorithms are fast, and they can make decisions based on far more information and factors than humans, consistently, and at scale.[27] Algorithms are, however, bad at ethics or following norms, do not justify their decisions, and are especially dependent on their training data and the data fed into models, which makes them prone to reproduce the biases in them. They are thus bad at edge cases.[28] Humans, on the contrary, are flexible decision-makers. We can exercise discretion, generalize and jump across context, even if actual decision-making processes are also opaque.[29] Hybrid systems thus promise to bring the best of both worlds by allocating tasks to either an individual or a machine, based on lists of what each is supposed to be better at.[30] To do so, the most popular methods construe humans and machines based on their capabilities and, on that basis, determining which capabilities can and should be automated and which ones shouldn't.[31] Thus, for example, Article 22 of the GDPR introduced a data protection right to "not be subject to a decision based solely on automated processing (…) which produces legal effects."[32]

[14] See Juan Ortiz Freuler, "Dataification, Identity, and the Reorganization of the Category Individual" (2023) 65 TLR.

[15] Bernard Marr, *Machine learning in Practice: How Does Amazon's Alexa Really Work?,* Bernard Marr & Co. (n.a.) available at: https://bernardmarr.com/machine-learning-in-practice-how-does-amazons-alexa-really-work/; *How Machine Learning is Used in Autonomous Vehicles,* Rinf.Tech (n.a.) available at: https://www.rinf.tech/how-machine-learning-is-used-in-autonomous-vehicles/#:~:text=An%20autonomous%20vehicle%20can%20use,the%20world%20around%20a%20car.

[16] ML is broadly defined as a methodology and set of data- driven techniques to come up with novel patterns and knowledge and to generate models that can be used for effective predictions about the data see Brent D. Mittelstadt and others, "The ethics of algorithms: Mapping the debate" (2016) 3 BD&S 2.

[17] See Mittelstadt (n16) 3.

[18] Mittelstadt (n16) 3.

[19] Jenna Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms" (2016) BD&S 3(1).

[20] See Donella H. Meadows, *Thinking in systems: A primer* (Chelsea Green Publishing, 2008).

[21] See Commission Report on safety and liability implications of AI, the Internet of Things and Robotics, 16 (Commission Report on safety and liability implications of AI) <https://commission.europa.eu/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics-0_en> accessed 26 August 2023.

[22] Burrell (n19).

[23] Expert Group on Liability and New Technologies, New Technologies Formation, Liability for Artificial Intelligence and other emerging digital technologies (2019) (Expert Group on Liability and New Technologies), 33; Helen Nissenbaum "Accountability in a computerized society," Science and engineering Ethics, 2(1) 29.

[24] Kaminski (n1).

[25] See H. James Wilson and Paul R. Daugherty, *Collaborative Intelligence: Humans and AI Are Joining Forces* (Harvard Business Review, July-August 2018) <https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces > accessed April 29 2024.

[26] In 2020 Jessica Fjield et al. found that out of 36 AI ethics documents, 70% included a principle proposing that important decisions made by AI systems be subject to human review see Jessica Jeld et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, (2020), https://papers.ssrn.com/abstract=3518482 (last visited Aug 27, 2023). Similarly, in 2021 Ben Green identified at least 41 policy documents from around the world that included some form human oversight requirement for algorithms in the public sector, including the AI Act. Ben Green, "The Flaws of Policies Requiring Human Oversight of Government Algorithms" (2022) CLSR 45.

[27] See Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, "Humans in the Loop" (2023) 76 VLR. 429, 464. Recent research is defining new types of interactions between humans and machine learning algorithms at the learning process see Eduardo Mosqueira-Rey and others, "Human-in-the-Loop Machine Learning: A State of the Art" (2022) 56 AIR 3005.

[28] Crootof and others (n27) 465.

[29] Crootof and others (n27) 462.

[30] Sidney Decker & David Woods, "MABA-MABA or Abracadabra? Progress on Human–Automation Co-Ordination," (2002) CTW4, 240.

[31] Joost de Winter and Dimitra Dodou, "Why the Fitts list has persisted throughout the history of function allocation" (2014) CTW16 (2014); Decker & Woods (n30) 104.

[32] Art. 22 GDPR.

This approach effectively precludes or restricts decision-making that is fully automated so that algorithmic predictions act more as an aid rather than a substitute to human decision-making.[33] Echoing this approach, the EU's Artificial Intelligence Act imposes an obligation for developers and deployers of high-risk systems to design and develop them so that they can be effectively overseen by a natural person.[34] Relying on AI human supervision, the proposed AI Liability Directive, which will be further discussed in Part 3, explains that "[t]here is no need to cover liability claims when the damage is caused by a human assessment followed by a human act or omission, while the AI system only provided information or advice which was taken into account by the relevant human actor."[35]

The optimism about leveraging human-machine interaction is problematic, however, and tempered by evidence that human-machine systems have dynamics of their own and it is difficult to design effective hybrid systems that require collaboration between humans and automated technologies.[36] This occurs for two main reasons: First, the assumption that people and computers have fixed strengths and weaknesses that can be easily capitalized on or used to compensate for the other party's weaknesses is not accurate.[37] Hybrid systems create new human strengths and weaknesses and it is a priori not obvious how to capitalize on different strengths.[38] For example, when automation can perform complex and repetitive tasks for an extended period, it increases the difficulty for humans to remain attentive and vigilant to the system. This can lead to a potential problem known as "vigilance decrement."[39]

Second, there are two competing tendencies in humans interacting with machines that have been observed: automation bias and algorithmic aversion.[40] Algorithmic aversion refers to the phenomenon of individuals wanting to override machine predictions even when these are highly reliable. Some of this originates from a perceived lack of agency, lack of transparency, and lack of trust in how accurate the system is.[41] Studies have thus shown that users will sometimes prefer to sacrifice accuracy for control over the algorithm's output.[42] Automation bias, on the other hand, refers to individuals' tendency to defer to automated systems even when they are wrong.[43] This can lead to a situation where the human does not detect problematic cases or fails to act even if they do; a famous example is pilots who tend to rely blindly on automated cues and don't remain vigilant.[44] Studies have shown that time pressure, complex tasks, and the degree of the user's self-confidence on their decisions tend to contribute to automation bias.[45] In other instances machines alone may be better at certain tasks.

A vast field of research has sought to identify ways to overcome some of the challenges of effective human-machine collaboration.[46] Much of this work has highlighted that since allocating a particular function to machines also creates new functions for humans, these must be accounted for in training and interaction. With other technologies these have included a transition to typing, or interacting with a screen and searching for the right display.[47] In the AI context, a move towards supporting better conditions for human-AI interaction requires making the operations of automated systems observable to humans and making it easy and efficient for human operators to direct the system, especially in novel episodes.[48] This should also be done paying special attention to the different levels of expertise, experience and training of the individuals interacting with these systems.[49] Maria de Arteaga and co-authors identify, for example, that supervisors of an algorithmic system in child-welfare in the US were able to correct for a glitch in the system because they had access to the underlying administrative data. This provided them with an alternative view of the case than what was being shown in the risk score calculation.[50] Other researchers have explored the idea of offering explanations for algorithmic decision-making[51] and incorporating forms of accountability to incentivize the reduction of automation bias.[52]

From the governance perspective, Crootof et al. have drawn from the experience of successful regulation of human-machine systems in safety-critical systems, to emphasize that hybrid human-AI systems require detailed rules for system designers and operators.[53] Regulation should require that product designers create technological systems around the people operating the system, that the devices are designed and labelled sufficiently for effective use, and address training and organizational policies.[54] Talia Gillis and her co-authors have relatedly highlighted the importance of taking into account the kind of interaction that is expected from humans and machines when designing these systems, but also, that oversight requirements be built that appropriately consider the combined and expected impact of the machine and human interaction and how it is implemented.[55] Indeed, substantive oversight requirements such as transparency or scrutiny of the data with which algorithms are trained seem to assume that the outcome that should be scrutinized and monitored is the algorithmic component of the decision in isolation. However, the true impact of AI systems is also the result of the human decision-making that

[33] Talia Gillis, *Regulating for "humans-in-the-loop"* (ECGI blog, 2022) <https://www.ecgi.global/publications/blog/regulating-for-humans-in-the-loop> accessed April 29, 2024; *see also* Tal Zarski, "Incompatible: The GDPR in the Age of Big Data Seton Hall Law Review" (2017) 47 Seton Hall Review 4(2), arguing in the early days of the GDPR that the requirements article 22 could be be sidestepped by inserting human intervention into the process.

[34] AI Act (n3) Article 14 see below Part 2.

[35] European Commission, Explanatory Memorandum, Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM/2022/496 final, 1 (Proposal AILD) Recital 15.

[36] Green (n26) citing Lisanne Bainbridge, "Ironies of automation" (1983) Automatica 19(6).

[37] Decker & Woods (n30).

[38] Decker & Woods (n30).

[39] Decker & Woods (n30).

[40] Maria De-Arteaga, Riccardo Fogliato & Alexandra Chouldechova, "A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores" (2020) Proceedings CHI Conference on Human Factors in Computing Systems 1.

[41] Kun Yu et al., Trust and Reliance Based on System Accuracy: 24th International Conference on User Modeling, Adaptation, and Personalization (2016) Proceedings UMAP 2016 223.

[42] Kun Yu et al. (n41).

[43] Green (n26).

[44] De-Arteaga and others (n40) citing Kathleen L Mosier and others "Automation Bias: Decision Making and Performance in High-Tech Cockpits," (1997) 8 IJAP 47.

[45] De-Arteaga and others (n40).

[46] Much of it developed in the past 30 years for aviation and surface transportation settings see Decker & Woods (n30).

[47] Decker & Woods (n30); Dale Richards and others, *"Designing for Human-Machine Teams: A Methodological Enquiry"* (2022) IEEE 3rd International Conference on Human-Machine Systems (ICHMS).

[48] Decker & Woods (n30).

[49] Crootof and others (n27) 498.

[50] De-Arteaga and others (n40).

[51] De-Arteaga and others (n40) 4.

[52] Linda J. Skitka and others, "Automation Bias and Errors: Are Crews Better than Individuals?" (2000) 10 IJAP 85.

[53] Crootof and others (n27) 494-496.

[54] Crootof and others (n27) 466; Green (n26) 14, emphasizing the importance of strengthening institutional oversight of algorithms, requiring justifications as to why it is appropriate to incorporate an algorithm into decision-making and to provide evidence that the algorithmic system can be effectively overseen.

[55] Gillis (n33).

accompanies it.[56]

In the last Part of this Article, I propose a related system-wide approach to AI liability. Before discussing the AI liability directive specifically, however, the next section taps into law and economics scholarship and presents the EU on AI governance and AI liability to propose a simplified framework of analysis for AI liability regimes.

2.3. The choice for regulation and liability for AI

Societies use two main institutional mechanisms to control the risks generated by new technologies like AI: Liability law and regulation. Liability law intervenes ex-post (only when harm occurs).[57] The objective is both to provide corrective justice and provide the right incentives to avoid harm.[58] Safety and risk regulation intervene ex-ante. Under these regimes, the government determines the optimal level of care for risk creators and seeks to modify behavior before and independently of the actual harm. It does so by prescribing, for example, specific technological or organizational requirements or that certain outcomes or processes be met.[59] The EU AI strategy uses both.[60]

Law and economics scholars have long studied when societies should recur to civil liability – whether strict liability or fault-based liability -, to regulation, or to both to control risks and harms. Steven Shavell identifies five main relevant factors to evaluate the desirability of any of these methods for controlling harm: quality of information of the state,[61] information available to victims,[62] the level of activity of the injurer,[63] the role of victims in diminishing harm,[64] and the administrative costs associated with enforcing liability or regulation.[65]

Most of the activities of everyday life are successfully regulated with civil liability: Many harms are easy to identify, and parties have the means and knowledge to mitigate harm at optimal levels (for example, what to do so that a tree in my property does not fall and affect my neighbor's property). Risks would be very difficult to address via regulation, as it would require frequent, intrusive, and expensive verification procedures.[66] Not by coincidence in most domestic regimes, the general rule for liability attribution is fault-based, which requires that the injurer's objectionable and avoidable conduct - fault - caused the damage.[67]

As scholars and policymakers have noted in the EU and elsewhere, this isn't necessarily the case with AI systems.[68] AI system's complexity, from a technical and organizational perspective – such as when humans and different organizations intervene in a particular outcome – complicate proving the key elements of fault-based liability. In the EU, the Expert Group on AI Liability, convenedby the Commission, identified in its influential 2019 report ("the Report") that regulation, such as product safety regulation, offer some safeguards to minimize the risks of harm when new technologies are rolled out in the market. It highlights those regulations, must be complemented with liability laws, some of which must be adapted, as they do not (and cannot) completely exclude the possibility that harm may occur.[69]

What follows briefly highlights the challenge for liability law in the EU, drawing mainly from the Expert Group's Report and presenting the case for, and a way to analyze, a regime that taps into the complementarity of regulation and AI liability.

a. The challenge for liability law

The characteristics of AI systems and their applications complicate the process of accessing compensation to victims of harm in all the cases where it seems justified. Additionally, the allocation of liability can be unfair or inefficient.[70] This occurs, for somewhat different reasons, in both of liability regimes in the EU: fault-based liability, and strict liability (which includes product liability).

In a fault-based liability regime, a victim of AI harm will face important obstacles establishing the three elements of fault-based liability: a harm, a wrongful action or omission by another person (fault), and causation. This can occur because (1) harm from certain types of actions may not be immediately obvious. This may be the case, of cases of AI bias in loans or subsidy applications where victims and other observers face difficulties knowing that a decision made by or with an AI system can be biased and can illegally discriminatory against them.[71] This "information gap", as Marta Ziosi and co-authors call it, can be

[56] Gillis (n33).

[57] European Group on Tort Law, Principles of European Tort Law, Art. 1:101 <http://egtl.org/docs/PETL.pdf > accessed 30 October 2023 (PETL).

[58] See Miriam Buiten, Alexandre de Streel and Martin Peitz, "The law and economics of AI liability" (2023) 48 CLSR p 4. Some authors also highlight that, since liability law creates incentives to take precautions ex-ante, it is also a form of risk regulation. Catherine M. Sharkey, presentation at "Free Expression and the DSA: Private-Public Workshop." Paris, France, Sciences Po Law School, June 10 and 11 2024.

[59] Marsden (n5) 1.

[60] See infra Part 4.

[61] Steven Shavell, "Liability for Accidents" in *Handbook of Law and Economics Vol. 1* (Eds. Mitchell Polinksy and Steven Shavell, 2007) 176 < http://www.law.harvard.edu/faculty/shavell/pdf/07-Shavell-Liability%20for%20Accidents-Hdbk%20of%20LE.pdf > accessed May 1, 2024.

[62] Shavell (n61) 176.

[63] Shavell (n61) 177.

[64] Shavell (n61) 177.

[65] Shavell (n61) 177.

[66] Steven Shavell, "A model of the optimal use of liability and safety regulation," (1984) 15 2 Rand Journal of Economics p 368.

[67] PETL (n57), Art. 1:101(2) a). Shavell explains that a fault-based liability is optimal when potential injurers have (1) enough information to know how to take care and the state has less information, so that it may err at determining what are optimal actions to prevent harm; (2) victims have a role to play in mitigating harm (by taking care as well); and (3) the administrative costs of verification and proving the elements of liability, harm included, is not excessively costly see Shavell (n61) 176.

[68] In 2017, for example, the Parliament adopted a Resolution urging the Commission to propose legislation on civil law rules for robotics and AI liability. In 2018, the Commission published a Staff Working Document on liability for emerging digital technologies which accompanied the Commission's Communication on Artificial Intelligence for Europe *See* Resolution on Civil Law Rules on Robotics, Eur. Parl. Doc. 2015/2103(INL) (2017), http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html. See also Laura Coppini, Robotica e intelligenza artificiale: questioni di responsabilità civile, 4 Politica Del Diritto 713 (2018); Commission Staff Working Document, Liability for emerging digital technologies, accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial intelligence for Europe, SWD/2018/137 final (2018).

[69] See Commission Report on safety and liability implications of AI (n21).

[70] Expert Group on Liability and New Technologies (n23), 1.

[71] In a well-documented scandal in the Netherlands an algorithmic decision-making system used by the tax authorities falsely accused tens of thousands of parents and caregivers. Yet, only in 2019, did it become apparent that the system was biased, while the system had been in place since 2013, even if victims maybe had a sense that something wrong was going on. Meilssa Heikkilä, "Dutch scandal serves as a warning for Europe over risks of using algorithms," (Politico.eu March 29, 2022) https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/?tpcc=nleyeonai; see also Marta Ziosi et al. "The EU Liability Directive (AILD): Bridging Information Gaps" (2024) European Journal of Law and Technology <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4470725>.

aggravated by the organizational opacity often surrounding AI systems.[72] (2) Proving fault is equality complicated given AI's opacity and complexity. It is hard to identify who was at fault, and a lack of behavioral standards complicates establishing what is the standard of care that different parties must follow.[73] This would require showing, for example, how others in the industry or field would have acted in similar circumstances and proving that the defendant's actions fell short of this expected standard, something that is hard to do given, in general, the opacity of the AI industry.[74] In the case of human-AI hybrid systems, the lack of clarity of how a particular system is supposed to improve human decision-making, and vice versa, creates additional difficulties in establishing to what extent the human in the loop contributed to harm or the contributing victim.[75] (3) Lastly, and for similar reasons, proving the cause-and-effect relationship between the defendant's actions or omissions and the resulting harm can be significantly hard. Given AI's technical and organizational opacity, doing things such as identifying how a bug in intricate software code, or the process behind an AI system's decision-making leads to a specific outcome, or gathering relevant evidence is more difficult, time-consuming, and expensive.[76]

Similarly, given current product liability law, the victims of AI harm will also face important challenges in succeeding at liability claims. Product liability law is usually understood as a form of strict liability, based on the principle that "the producer" of a product is liable for damages to life, health and property caused by a defect in a product they have put into the market as part of their business regardless of whether the defect is their fault.[77] Some scholars have highlighted, that the definition of a defect as something that could have been known at the time of placing a product in the market makes it more similar to fault-based liability.[77] In any case, European authorities and the Expert Group for AI Liability have identified that the Product Liability Directive of 1986 (PLD) regime is not fit to meet the risks of emergent technologies like AI: This occurs because systems challenge the notions of a "product" and a "defect."[78] The PLD (1) only covers tangible products, which included software and AI integrated into tangible products, but not to standalone software products.[79] (2) Defectiveness is determined based on the safety expectations of the average consumer, but so long as the defects could have been known at the time the product was placed on the market. (3) The PLD focuses on the moment when a product was put into circulation as the moment that defines the producer's liability, this cuts off claims over subsequent additions – by the producer or someone else – over updates or upgrades or a system. It also does not account for software updates which are often meant to make products safer but users may not install, or the fact that AI systems are supposed to continue learning once they are placed in the market nor does it provide duties to monitor products after circulation.[80]

From a policy, and law and economics perspective, an alternative would be to extend a "stricter" version of strict liability to AI producers, regardless of who is in control and regardless of whether the defect was known. Indeed, particularly from the 19th century onwards, legislators often responded to risks brought about by new technologies - like trains and motor vehicles - by introducing strict liability, a liability regime that does not require the injurer's conduct to have been faulty but merely that their conduct caused harm.[81] Professor Christiane Wendehorst has recommended, for example, that a harmonized regime of vicarious liability be adopted so that "a principal that employs AI for a sophisticated task faces the same liability under existing Member State law as a principal that employs a human auxiliary."[82] This would address the difficulty victims have in proving fault or defectiveness. Legislators and courts would not need to have information on the optimal level of precaution in designing and deploying AI-based systems.[83]

Law and economics scholars highlight that, indeed, strict liability creates optimal incentives to reduce socially wasteful accidents: By removing the fault requirement, strict liability creates incentives for care where a potential injurer would find that it cheaper, under a fault regime, to eventually pay for damages than to prevent damage.[84] Additionally, strict liability is easier to prove (one factor, negligence, does not have to be proven).[85] Importantly, strict liability also directly induces changes in the levels of the activity as issue as higher levels of activity increase the likelihood of harm, regardless of level of care. Thus, if a potential injurer believes that they can achieve the same business results, but at lower activity levels, they will do so.[86]

As the Expert Group and other scholars have noted, however, strict liability is less useful in cases when the AI systems are complex and there is a human in the loop: strict liability for producers would not create enough incentives for AI operators nor victims to take optimal precautions.[87] Indeed, in instances where harm can also be avoided by encouraging changes in activity by victims or other actors, law and economics scholars don't encourage strict liability either.[88]

The Expert Group also notes that strict liability may have important impacts on technological advancement. Some individuals or entities may become more hesitant to actively promote technological research if the risk of liability is perceived as a deterrent.[89] Activities that are beneficial to society but also risky may be reduced below the optimal level because costs will be internalized while positive externalities will not flow back directly to developers, even when sufficient precautions are in place.[90] This could be the case in instances where AI's exceptional performance reduces harm to society compared to not using AI at all –

[72] Amnesty International, "Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal," (Amnesty International, 2021) https://www.amnesty.org/en/documents/eur35/4686/2021/en/ accessed October 30, 2023, 12, 15; Ziosi (n71).

[73] See Buiten and others (n58) 7; Expert Group on Liability and New Technologies (n23) 20; *see* also discussion of the AI Act above.

[74] See Expert Group on Liability and New Technologies (n13) 26; Buiten and others (n58) argue that, in the case of autonomous AI systems, this is aggravated by the fact that some outputs can't be anticipated. This challenge may be mitigated however, upon interacting with other risk-mitigating regulations where AI systems are specifically trained to avoid certain harmful outputs).

[75] Expert Group on Liability and New Technologies (n23) 31.

[76] Expert Group on Liability and New Technologies (n23) 26.

[77] Directive (EU) 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products, Art. 4 and 7 (Product Liability Directive). See also Richard Posner, "Economic Analysis of Law" p. 165. 3rd Edition (1986) at 166 arguing that "the term strict liability is something of a misnomer here, because in deciding whether a product is detective or unreasonably dangerous in design or manufacture the courts often use a Hand Formula approach, balancing expected accident costs against the costs of making the product safer."

[78] Expert Group on Liability and New Technologies (n23) 30.

[79] Expert Group on Liability and New Technologies (n23) 19.

[80] Expert Group on Liability and New Technologies (n23) 30.

[81] Miquel Martín-Casals, *Technological Change and the Development of Liability for Fault: A General Introduction,* The Development of Liability in Relation to Technological Change (Miquel Martín-casals, et al. eds. 2010).

[82] Christiane Wendehorst, *Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks*, The Cambridge Handbook of Responsible Artificial Intelligence (Silja Voeneky et al., eds. 2022), 208.

[83] Wendehorst (n82).

[84] Posner (n77) 160.

[85] Posner (n77) 164 Shavell adds that an outcome where victims are not encouraged to take due care, where they could, is also inefficient see Steven Shavell "Strict Liability versus Negligence," The Journal of Legal Studies, 7.

[86] Posner (n77) 161.

[87] See Expert Group on Liability and New Technologies (n23); Buiten and others (n58) at 10.

[88] Posner (n77) 162.

[89] Expert Group on Liability and New Technologies (n23), 28.

[90] Expert Group on Liability and New Technologies (n23), 10.

such as AI diagnostic tools that outperform humans in disease detection. This is also below their optimal level.[91] While the use of AI reduces harm when compared to other options, there are also opportunity costs associated with not utilizing AI.[92]

b. The place of regulation and the joint use of liability law

Regulation is well suited to control harm in instances where there are sufficiently important factors that dilute the incentive to take care under liability. This is the case when the regulatory entity has an information advantage, or where it may be desirable to compel parties to produce information that they do not produce ;[93] where the potential harm is very large and would exceed companies' capacity to compensate harmed people,[94] and where responsible parties perceive that they may not be sued in case of harm.[95] This can occur, for example, when harms are dispersed and individual victims may not find it worth it to suit, when harms are hard to identify and/or only become apparent later on, and where it is difficult to trace the harm to particular causes or firms.[96] In these cases harmed individuals will find it hard, or not cost-effective, to sue and show the main requirements of liability.

The choice for these harm-control tools is not, however, exclusive. Rather, regulation and civil liability can complement each other well in some instances. In France, for example, Pierre Bentata found that in the management of hazardous operations, judges and regulators interact in interesting ways and often provide each other with important information: the number of cases increased sharply after regulations were passed and victims' chance of success seems to increase. Additionally, Bentata observes that most of the plaintiffs are the most heavily regulated facilities and the state-owned companies and that judges are more severe against the latter. This, he suggests, appears to be a way in which civil liability reduces risks of regulatory capture, and offers an additional level of deterrence that goes beyond the one offered by regulation.[97]

Relatedly, Shavell also found that liability and regulation can be designed so that they complement each other optimally. He finds that ultimately neither regulation nor liability alone led all parties to exercise the socially desirable standard of care. This occurs for the reasons already highlighted above: regulatory authority's information about risk is often imperfect (and so it will sometimes ere setting the right standard), and because liability will sometimes not create sufficient incentives to take sufficient care (because they may not be sued for it, for example).[98] Shavell explains that it may be thus advantageous to use both tools so that they have the following effects: Regulation sets a baseline for all parties covered by it to take a certain level of care. However, parties causing more than relatively low risks will be led to do more than is required by the regulatory standard, because they will be further deterred by the likelihood of being held liable. Regulators can also reduce the standard of regulation, and thus reduce the cost of compliance, since liability compensates for some of the slack associated with the lower standard but is based on parties' better information.[99]

c. Towards a mixed use of AI regulation and liability to control AI risks

The question guiding this Article is how policymaker draft AI liability should complement AI regulatory frameworks considering AI complexity and the plurality of actors and people that participate in AI systems. It does focus on how the AI liability directives proposed by the EU Commission are set to complement the AI Act.

Drawing from the analysis and review conducted in this section there seem to be two key factors that should be examined when analyzing how the proposed liability rules complement the AI Act: First, in the case of harm, the AI liability framework makes it easy for victims to bring a liability claim against AI producers or deployers. Second, given the level of activity chosen by the AI Act, the liability framework is capable of encouraging AI developers and deployers who create more than low risk to take more care.

3. The AILD and PLD in the context of the European AI strategy

The EU is a world leader in AI governance.[100] The EU AI strategy, first announced in 2017, seeks to establish a general EU-wide coordinated approach "to make the most of the opportunities offered by AI and to address the new challenges that it brings."[101] At the regulatory level it seeks to establish an appropriate ethical and legal framework that would support "an environment of trust and accountability around the development and use of AI."[102] Three interrelated legal initiatives seek to create the ecosystem of trust sought by the Commission: The AI Act, approved in 2024, seeks to address fundamental rights and safety risks; a civil liability framework, which is composed of the directives at issue here, the revision of the PLD and a the AILD; and a revision of sectoral safety legislation, such as Machinery Regulation and the General Product Safety Regulation.[103] (This piece does not discuss directly the relevant sectoral safety regulations).[104] At the time of writing the liability framework for AI systems is under consideration in the EU parliament.

This Part briefly presents the two proposed AI liability directives as they relate to the framework set in place by the AI Act.

3.1. The AI Act

The cornerstone of European AI regulation is the AI Act. This Act is an umbrella and union-wide framework adopting a risk-based approach to AI regulation to ensure embedded safety and security in products and services.[105] It aims to promote human-centric and trustworthy AI while safeguarding health, safety, fundamental rights, democracy, the rule of

[91] This may be the case of some instances in medicine and some autopilots, like airplanes.

[92] Buiten and others (n58) at 10 discussing from a law & economics perspective how "the chosen liability regime should therefore be seen in the context of public policy towards innovation."

[93] Shavell (n66) at 361.

[94] Shavell (n66) at 369.

[95] Shavell (n66) at 370.

[96] Shavell (n66) 370.

[97] Pierre Bentata, (2014) "Liability as a Complement to Environmental Regulation: An Empirical Study of The French Legal System," Environmental Economics and Policy Studies, vol. 16, 722.

[98] Shavell (n66) 271.

[99] Shavell (n66) 272.

[100] See Anu Bradford, *Digital Empires. The Global Battle to Regulate Technology* (Oxford University Press, 2023).

[101] European Commission, Communication from the Commission, Artificial Intelligence for Europe, COM(2018) 237 final (Communication Artificial Intelligence for Europe).

[102] Communication Artificial Intelligence for Europe (n102).

[103] European Commission, "A European approach to artificial intelligence", Shaping Europe's Digital Future, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

[104] The proposed Machinery Regulation and the proposed General Product Safety Regulation which revise the existing Machinery Directive and General Product Safety Directive, aim, in their respective fields, to address the risks of digitalization in product safety, but not liability. *See* European Commission, "The General Product Safety Directive" <https://commission.europa.eu/business-economy-euro/product-safety-and-requirements/product-safety/consumer-product-safety_en> accessed October 30, 2023. European Commission, "Machinery"<https://single-market-economy.ec.europa.eu/sectors/mechanical-engineering/machinery_en> accessed October 30, 2023.

[105] Commission Report on safety and liability implications of AI (n21) 4.

law, and the environment from AI's harmful effects, all while fostering innovation.[106] What follows is a concise overview of the main safety requirements introduced by the AI Act, with a focus on human oversight and the importance of standardization and conformity assessments in the AI's implementation.

a. Levels of risk and key safety requirements

The AI Act applies to providers and deployers of AI systems in the EU. The Act defines providers as the natural or legal person who develops an AI system with a view of placing it in the market, and deployers as the natural or legal person that uses the AI system.[107] It categorizes AI systems into four risk levels based on their intended use and regulates them differently, banning systems that pose certain unacceptable risks, and imposing certain requirements on the rest.[108] Most of the Act is concerned with the safety requirements for high-risk systems, identified in Annex III. Late in the process of passing the Act, the EU parliament introduced amendments for providers of general-purpose AI models in response to the emergence of generative AI.[109] General-purpose AI models are classified as with systemic risk when it has high-impact capabilities, based on their computational power and indicators and benchmarks still to be defined.[110] They have similar requirements as high-risk systems.[111] Limited risk systems must comply with minimal transparency requirements to enable informed user interaction.[112]

Providers of high-risk systems must comply with the seven key requirements: (1) Implementing and maintaining a risk management system throughout an AI system's lifecycle.[113] (2) Evaluating the availability, quantity and suitability of the data used for training models, identifying biases and gaps that need to be addressed.[114] (3) Creating and updating technical documentation of high-risk systems before they are placed on the market.[115] (4) Designing AI systems to automatically record operational events to ensure that the AI system's functioning is traceable.[116] (5) Designing AI systems so that their operation is "sufficiently transparent to enable deployers to interpret the system's output and use it appropriately."[117] AI systems must also be accompanied by instructions for use, and human oversight measures to facilitate the interpretation of AI outputs.[118] (6) Designing and developing high-risk AI systems so that they enable effective human oversight while in use.[119] (7) AI systems shall be designed and developed to achieve an appropriate level of accuracy, robustness and cybersecurity.[120]

Distributors, importers, deployers and other third parties will be considered providers when they put their name or trademark on a system on the market when they make substantial modifications to it, or when they modify their intended purpose.[121] Providers can demonstrate conformity with these requirements through self-assessment and internal control. If they conform with harmonized standards, they will be presumed to be compliant with the requirements of the Act and EU law protecting fundamental rights.[122]

b. The human-in-the-loop requirement

One of the key objectives of the EU's regulatory framework is to promote the development of AI systems that function "in a way that can be appropriately controlled and overseen by humans."[123] Early versions of the AI Act were criticized for their reliance on the human-in-the-loop as safety requirements.[124] The latest version at the time of writing seems to have tried to accommodate some of the research, and critiques to human-in-the-loop requirements presented in Part I C. Attempting to accommodate the concerns presented in Part I C, the current version of the Article emphasizes the design of "appropriate human-machine interface tools" so that high-risk AI systems can be "effectively overseen by natural persons."[125] It also requires that individuals in charge of the oversight must have sufficient AI literacy, and are appropriately enabled to understand and interpret the system, be aware of the possibility of over-relying on the system be able "to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system; [and be able] to intervene on the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure."[126]

c. Reception and critique of the AI Act

The reception of the AI Act in the EU was mixed.[127] European institutions saw it as a major success, positioning the EU's leadership as

[106] AI Act (n3).

[107] AI Act (n3) Article 3(3). Note that individuals who are subject to AI systems have no role to play in the AI act. This is according to the latest version of the Act. In former versions, the Act has referred to deployers, as users.

[108] European Parliament, "EU AI Act: first regulation on artificial intelligence" (News-European Parliament, June 14, 2023) <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> accessed 25 April 2024.

[109] See Dr. Benedikt Kohn and Lennart Van Neerven, Will Disagreement Over Foundation Models Put the EU AI Act at Risk? (*Tech Policy Press*, November 29, 2023), <https://www.techpolicy.press/will-disagreement-over-foundation-models-put-the-eu-ai-act-at-risk/> accessed May 2, 2024.

[110] AI Act (n3) Article 51.

[111] European Parliament (n108).

[112] European Parliament (n108).

[113] AI Act (n3) Article 9.

[114] AI Act (n3) Article 10.2.

[115] AI Act (n3) Article11.

[116] AI Act (n3) Article 12.

[117] AI Act (n3) Article 13.1.

[118] AI Act (n3) Article 13.2, Art 13.3.

[119] AI Act (n3) Article 14.

[120] AI Act (n3) Article 15.

[121] AI Act, (n114) Article 25.

[122] AI Act (n3) Article 40.

[123] AI Act (n3) Recital 27.

[124] Crootof and others (n27), Green (n26).

[125] AI Act (n3) Article 14(1).

[126] AI Act (n3) Article 14(4).

[127] Emma Woolacott, "European Union's AI Act Gets Mixed Reception" (*Forbes*, March 19, 2024) < https://www.forbes.com/sites/emmawoollacott/2024/03/19/eus-ai-act-gets-mixed-reception/?sh=5145a22042c9 > accessed May 2, 2024.

the global leader in technology regulation.[128] Certain think tanks and companies worry that the new rules may overburden innovative AI developers, and highlight that a lot of uncertainty remains about the implementation of the Act.[129] Human and digital rights activists argue that the Act did not go far enough to protect individuals from AI harms.[130] Indeed, the tiered-risk approach inevitably leaves out certain applications that may be risky. Critiques highlight, however, that the last version highlights that the final version of the Act seems to have been particularly lenient to particularly sensitive realms such as national security (where a blank exception was adopted), and by allowing uses of facial recognition and other biometric categorization systems by law enforcement and migration authorities, while they are prohibited in education and the workplace.[131]

Another critique is that while the latest version of the Act provides some individual remedies, like lodging complaints or receiving explanations for decisions, it lacks robust rights and redress mechanisms.[132] The AI liability regime, however, is supposed to offer those mechanisms of redress.[133] The next sections briefly present the two directives, and the last Part evaluates how they complement the AI Act.

3.2. The revised PLD: liability for "material damages caused to natural persons by AI-powered products"

The revision of the Product Liability Directive seeks to adapt the EU's product liability regime to new technologies.[134] The Revised PLD aims to ensure that liability rules reflect the nature and risks of the new digitally powered products, easing the burden of proof in complex cases and easing restrictions on making claims "while ensuring a fair balance between the legitimate interests of manufacturers, injured persons and consumers in general."[135] As its predecessor, the proposed directive establishes a form of strict liability of the relevant economic operators "as the sole means of adequately solving the problem of a fair apportionment of the risks inherent in modern technological production."[136] Claimant must prove the elements of product-strict liability: defectiveness of the product, the damage suffered and the causal link between the defectiveness and the damage.[137] What follows describe the main changes and how they apply to AI systems:

Chapter 1 lays out the subject matter, the scope of the directive, and key definitions. Importantly, the Revised PLD changes the definition of product and economic operators to extend it to software, AI systems and AI-enabled goods, such as smart-home devices.[138] It also expands its application to digital services that are integrated or integrated with a product "in a way that would prevent the product from performing one of its functions,"[139] such as navigation software in an autonomous vehicle.[140] Additionally, it defines economic operators as the manufacturers of a product or a component, the provider of a service, and the importer or the distributors,[141] and extends liability to natural or legal persons that modify a product substantially after it has already been placed in the market will also be considered economic operators.[142]

Chapter 2 lays out the key rights and obligations of the product liability regime.[143] Defectiveness is defined as the circumstances when a product "does not provide the safety which the public at large is entitled to expect." This is to be determined considering the presentation of the product including instructions for installation and maintenance,[144] and the expectations of the end-users for whom the product is intended,[145] reasonable use and misuse of the product,[146] the safety requirements of the product,[147] the moment in time when the product was placed in the market and, importantly, the moment in time when the product leaves the control of the manufacturer.[148] The distinction between the moment in time at which a product is placed in the market, and the moment at which it leaves the manufacturer's control seeks to reflect that many products, such as AI systems, remain within the manufacturer's control even after being placed in the market.[149]

The Revised PLD also establishes a rebuttable presumption of defectiveness to alleviate the claimant's burden of proof.[150] This will be the case if the claimant establishes that the product does not comply with mandatory safety requirements or when the damage was caused by an obvious malfunction of the product during normal use and circumstances.[151] Additionally, the directive establishes that national courts must be empowered to order the defendant to disclose relevant evidence that is at its disposal, upon request of an injured person claiming compensation for damage caused by a defective product, and when the claimant has presented facts and evidence sufficient to support the

[128] Gian Volpicelli, "European lawmakers rubberstamp EU's AI rulebook," (Politico, March 13, 2024) <https://www.politico.eu/article/european-lawmakers-rubber-stamp-eus-ai-rulebook/> accessed May 2, 2024.

[129] Eliza Gkritsi, "The long and winding road to implement the AI Act" (Euractiv, March 14, 2024) https://www.euractiv.com/section/digital/news/the-long-and-winding-road-to-implement-the-ai-act/> accessed May 2, 2024.

[130] Gkritsi (n129).

[131] EDRi and coalition partners, "EU's AI Act fails to set gold standard for human rights" (*EDRi.org,* Arpil 3, 2024) < https://edri.org/our-work/eu-ai-act-fails-to-set-gold-standard-for-human-rights/> accessed May 2, 2024.

[132] EDRi and coalition partners (n131).

[133] See White Paper on AI (n8) and the discussion on the AILD below.

[134] See infra section 3.1(a) for why it needs updating.

[135] Expert Group on Liability and New Technologies (n23), 2.

[136] European Commission, Proposal for a Directive of the European Parliament, and the Council on liability for defective products, COM/2022/495 final, Recital 2 (Proposal New PLD).

[137] Proposal New PLD (n136) Article 9.

[138] Proposal New PLD (n136) Recital 12, see also Explanatory Memorandum New PLD (n125), 3 Art. 4(1).

[139] Proposal New PLD (n136) Article 4(4).

[140] Proposal New PLD (n136) Recital 15.

[141] Proposal New PLD (n136) Article 4(16).

[142] Proposal New PLD (n136) Article 7(4).

[143] Proposed New PLD (n136) Article 5.

[144] Proposal New PLD (n136) Art 6(a).

[145] Proposal New PLD (n136) Art 6(h).

[146] Proposal New PLD (n136) Art 6(b).

[147] Proposal New PLD (n136) Article 6(f).

[148] Proposal New PLD (n136) Art 6(e).

[149] Proposal New PLD (n136) Recitals, 22, 23.

[150] Proposal New PLD (n136) Recital 33.

[151] Proposal New PLD (n136) Article 9.

plausibility of the claim for compensation.[152] Defectiveness will be also presumed when the defendant fails to comply with an order to disclose relevant evidence,[153] and when it is established that the product is defective and the damage caused is of a kind typically consistent with the defect in question.[154]

Chapter 3 covers other general provisions on liability., Manufacturers and distributors will not be liable if they can prove that it is probable that the defect that caused the damage did not exist when the product was placed on the market or put into service ;[155] that the defectiveness is due to compliance of the product with mandatory regulations ;[156] that the product is up to the state of the scientific and technical knowledge at the time it was placed in the market.[157] However, economic operators will not be exempted from liability when the defect is due to software updates or upgrades, or a lack thereof.[158] Lastly, the proposed directive establishes that economic operators cannot reduce their liability when a third party's actions or omissions contributed to the harm.[159] In any case, in the interests of a fair apportionment of risk, when the damage was caused by the defectiveness of the product and the faulty action of a third party or the victim, their liability may be reduced.[160]

3.3. *The AILD: "Adapting non-contractual civil liability rules to artificial intelligence"*

The AILD seeks to adapt, in general, national liability rules to the challenges posed by claims for damages caused by AI-enabled products and services. It does so by laying out rules on the disclosure of evidence, and by establishing a rebuttable presumption of causal link in the case of fault.[161] By doing so, the AILD seeks to address the challenges that victims may face when an AI system participates in the action that led to the harm.[162] Interestingly, the directive explicitly does not adopt a stringer standard than fault-based liability (such as reversal of the burden of proof, or an irrebuttable presumption) because of how costly this could be for developers or deployers.[163] It is thus mostly oriented at ensuring ensures that victims of damage caused by AI have an equivalent level of protection under fault-based liability rules as victims of equivalent harms caused without AI systems.[164]

The directive is rather short: it has only nine Articles, four out of nine (Articles 5 to 9) which are concerned with the creation of a monitoring program to provide the European Commission with information on incidents involving AI systems and the implementation of the Directive in Member States.[165] Article 1 establishes its subject matter, Article 2, covers key definitions, mostly referring to the AI Act.[166] Articles 3 and 4 contain the key measures: rules for the disclosure of evidence; and conditions to establish a rebuttable presumption and a rebuttable presumption of the causal link between fault and harm.

The rules for the disclosure of evidence are in Article 3. In a nutshell, national courts must be empowered to demand the disclosure of relevant evidence from high-risk systems suspected of causing damage to providers or those subject to their obligations. This disclosure must strictly adhere to what is necessary and proportionate to support the claim.[167]

Article 4 lays the requirements for national courts to establish a rebuttable presumption of a causal link between the fault and the output of the AI system. National courts shall presume fault where three conditions are met: fault has been established or presumed according to Article 3, it can be considered likely that the fault influenced the output, and the claimant showed that the output led to the damage.[168] The causal link between fault and output will also be presumed when the claimant shows that the deployer of a high-risk AI system did not comply with its obligations under the AI Act.[169] Similarly, the presumption will be established when it is deployers who do not comply with their obligations to use or monitor the AI system following the accompanying instructions of use,[170] or if the claimant proves that the deployer "exposed the AI system to input data under its control which is not relevant given the system's intended purpose."[171]

For non-high-risk systems, the presumption of causality will apply only if the court determines that it is excessively difficult for the claimant to prove the causal link between damage and fault. This should be assessed given the characteristics of certain AI systems, such as their autonomy or opacity.[172]

Importantly, Recital 15 states that the AILD need not cover situations "when the damage is caused by a human assessment followed by a human act or omission, while the AI system only provided information or advice which was taken into account by the relevant human actor."[173] This is the case because, supposedly, when the damage is caused by a human assessment, "while the AI system only provides information or advice" it will be possible to trace back the damage to a human act or omission, and therefore establishing causality will not be as hard as

[152] Proposal New PLD (n136) Article 8.

[153] Proposal New PLD (n136) Article 9.

[154] Proposal New PLD (n136) Article 9.

[155] Proposal New PLD (n136) Art 10(c).

[156] Proposal New PLD (n136) Article 10(d).

[157] Proposal New PLD (n136) Article 10(e).

[158] Proposal New PLD (n136) Article 10.2.

[159] Proposal New PLD (n136) Article 12.1.

[160] Proposal New PLD (n136) Art 12.2 This echoes the principle of the contributory conduct tor activity of the victim *see* PETL (n2) Article 8:101, Recital 36.

[161] Proposal AILD (n35) Article 1(b), Article 4.

[162] See infra 3.1(a); Proposal AILD (n35).

[163] European Commission, Explanatory Memorandum, Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence, Explanatory Memorandum, COM/2022/496 final (Explanatory Memorandum AILD).

[164] Explanatory Memorandum AILD (n163), 10, see also Proposed AILD (n35) Article 1.

[165] Proposal AILD (n35) Article 5.

[166] Proposal AILD (n35) Article 2.

[167] Proposal AILD (n35) Article 3.1 paragraph 2.

[168] Proposal AILD (n35) Article 4.1.

[169] Proposal AILD (n35) Article 4.2.

[170] Proposal AILD (n35) Article 4.3(a).

[171] Proposal AILD (n35) Article 4.3(b).

[172] Proposal AILD (n35) Article 4.5, Explanatory Memorandum AILD (n163).

[173] Proposal AILD (n35) Recital 15.

when an AI system is involved. As other commentators have noted, this may leave significant amounts of the AILD proposal inapplicable, as the AI Act will require that high-risk systems be designed and developed so that they *can* be effectively overseen by natural persons (as proposed in the text by the Commission) or so that "they be effectively overseen by natural persons" (as proposed by the Parliament).[174]

4. Analysis of the EU AI liability regime

So, how does the proposed AI liability regime complement the AI Act in incentivizing precautionary measures and reducing socially wasteful AI accidents, considering the complexity of AI and the involvement of multiple actors?

This Part answers this question by first presenting three hypothetical accidents involving an AI system. Analyzing these three examples and using the framework presented in Part 2 it then concludes that the proposed directives make important progress in addressing the challenges AI systems pose to accountability and enhance the incentives for AI deployers, developers and users to take better care and avoid harm. The analysis shows too, however, that the current proposals fall short on two main accounts:

First, the AI Act's tiered risk regulation drifts over into the liability proposals, as it will be mostly in cases involving high-risk systems where victims will have better access to information and where most of the presumptions will apply. This leads to the regime being still not very effective at addressing the liability challenges for systems that are non-high risk but still complex and opaque.

Second, the regime's treatment of human-AI hybrid systems is still somewhat simplistic. Since AILD *excludes* from its application systems where AI is only advising humans but not effectively deciding, it may also create incentives for AI designers to design systems to "advise" humans, even if more collaborative or even entirely automated systems may be safer and better. This contradicts not only the research about better human-AI design and interaction, but also seems to contradict the final version of the AI Act, which incorporated some of the critiques on the challenges of effective human supervision and emphasizes the importance of effective design, instructions, and a turn towards human-AI collaboration and not, merely, supervision.

A last observation that follows from the analysis is that enforcing the AI liability regime will potentially – and maybe unavoidably – rely on the development of the technical standards mandated by the AI Act. Though liability standards of care – referring to the model of careful and prudent conduct required from the perpetrator of the damage – are in principle different from standards of quality and safety required by law and established standard-setting bodies, certain legal and technical standards will play a significant role in determining what is reasonable to expect from the various parties involved.[175]

4.1. AI and safety when a human is involved: the case of an autopilot

Imagine an accident involving a vehicle with an autopilot feature. This happens in a part of a city where using autopilot is allowed. Assume the AI Act, the PLD and the AILD are in place (as they were presented in the previous section), and that these are the main EU-law institutions that apply; there are no special liability nor product safety rules for Automated vehicles.[176] The vehicle swerved into a curb, causing the car accident which resulted in an injury to the driver. The car manufacturer cautions drivers to keep their hands on the wheel, and "be prepared to take over at any moment."[177] In the accident, the driver received a warning to control the vehicle less than a second before the strike, as this was when the software identified it was facing an unknown situation. The manufacturer says the software worked correctly.[178]

The driver sues the car manufacturer, alleging that the Autopilot feature failed to operate safely and caused the accident. In real-life cases like this, juries in the US have found that the Autopilot feature had not malfunctioned and that the driver's negligence caused the accident.[179] The legal issue in this case would thus be: *given the EU's new liability rules, is the manufacturer of an automated vehicle liable for an accident involving the AV, where the driver received a warning to control the vehicle less than a second before the strike, but where the car-manufacturer also warns drivers to be ready to take control anytime?*

From the victim's perspective, a good result would be that the manufacturer is found to be at fault, or that the software is found to be defective because it passed the control to the driver less than one second before the strike. From a societal perspective, if the driver is shown to have been distracted, a better result would be that both the driver and the software share some of the responsibility so that all parties have incentives to take optimal precautions in the future.[180]

The EU is expected to draft specific security rules for AVs, and they will be exempt from the core obligations of the AI Act. Let's assume, for this example, that these will be equivalent to those of the AI Act.[181] Because this is a claim about a bodily injury, suffered by a natural person, and caused by a product, let's also assume that this claim falls under the jurisdiction of the Revised PLD.[182] This is already beneficial for the plaintiff (although not new) as they would not have to establish

[174] See Philip Hacker, "The European AI Liability Directives – Critique of a Half-Hearted Approach and Lessons for the Future" Working Paper, at 19 <https://arxiv.org/pdf/2211.13960.pdf > accessed 30 October 2023.

[175] See e.g. Bryan H. Choi "NIST's Software Un-Standards" (2024) The Digital Social Contract: A Lawfare Paper Series < https://www.lawfaremedia.org/article/nist's-software-un-standards> accessed May 2 2024, (discussing how in the US, NIST's cyber frameworks are being invoked as standard of care and raising the question on whether they are adequate).

[176] Special liability rules for road accidents exist in several countries, which are commonly strict liability rules, as do special safety regulations for AVs. See David Fernandez Llorca and Emilia Gomez Gutierrez, "Artificial Intelligence in Autonomous Vehicles towards trustworthy systems", European Commission 2022 (JRC128170) <https://publications.jrc.ec.europa.eu/repository/handle/JRC128170> accessed October 30, 2023.

[177] This is, in fact, what drivers of Tesla's are expected to do Mike Spector, Dan Levine & Mike Spector, "Exclusive: Tesla Faces U.S. Criminal Probe over Self-Driving Claims" (*Reuters*, Oct. 27, 2022) https://www.reuters.com/legal/exclusive-tesla-faces-us-criminal-probe-over-self-driving-claims-sources-2022-10-26/ accessed August 25, 2023.

[178] This has happened in Tesla-related accidents Abhirup Roy, Dan Levine & Hyunjoo Jin, "Tesla Wins Bellwether Trial over Autopilot Car Crash" (*Reuters*, Apr. 22, 2023) <https://www.reuters.com/legal/us-jury-set-decide-test-case-tesla-autopilot-crash-2023-04-21/> accessed August 25, 2023.

[179] See Andrew J. Hawkins, "The world's first robot car death was the result of human error – and it can happen again" (*The Verge*, 20 November 2019) <https://www.theverge.com/2019/11/20/20973971/uber-self-driving-car-crash-investigation-human-error-results > accessed 30 October 2023.

[180] See Part 2; Buiten and others (n58).

[181] See Hacker (n174) 2: "Technically, autonomous vehicles will be considered high-risk (Article 6(1) and (2) AI Act) but are exempt from all of the core obligations of the AI Act (Articles 2(2) and 84 and Annex II Section B No. 2, 3, 6 and 7 AI Act), hence rendering the relevant references in Articles 3 and 4 AILD Proposal inapplicable to them."

[182] Proposal New PLD (n136) Article 1.

fault, they only have to prove that the product was defective, the damage suffered, and the causal link amongst both.[183] A second legal element is that the AV software is the high-risk category under the AI Act.[184] Thus, the AV manufacturer is obliged to meet safety requirements such as producing technical documentation and record keeping, and designing for human oversight, and transparency.[185]

According to Article 6 of the Revised PLD, a product is defective if it does not "provide the safety which the public at large is entitled to expect."[186] This includes the presentation of the product, instructions, etc. ;[187] the reasonably foreseeable use and misuse of the product,[188] product safety requirements,[189] and the specific expectations of the end-users for whom the product is intended.[190] Article 9, the defectiveness is presumed if, the plaintiff shows that the vehicle (1) does not comply with mandatory safety requirements of the product, or (2) that the damage was caused by an obvious malfunction.[191] Producers are exempt if the defect did not exist when the product was placed on the market,[192] if the defect is caused due to compliance of the product with mandatory regulations,[193] or if the state of scientific-technical knowledge at the time the product was placed on the market was not such that the defect could be discovered.[194]

Following the research on the complexities of AI-Human interactions, one of the questions such a case raises is whether handing over control less than a second before the accident is "the kind of safety the public at large expects," or according to the expectations of end-users who, in this case, is a regular driver (but not a professional car-racing driver, for example). If it isn't it would be a defect.[195] Indeed, handing over control in such a way seems like the kind of problematic interface inspired by the idea that humans and machines complement each other easily, discussed in Section 2.2.

To show that there is a defect, and with the PLD in place, the plaintiff would be able to request documentation and evidence from the vehicle's manufacturer about the system and its design.[196] In this case, this would include information on the technical documentation on the autopilot, the AI-Human interface but also, if this were a device covered by the AI Act, the conformity assessments with the requirements of the AI Act and, in general, what the expected duty of care of the producer is about human oversight.[197] Though there are, still to date, no such clear behavioural standards, the AI Act (or the future requirements for AVs in particular), may offer some guidelines about how this looks like at the design stage: There must be "appropriate human-machine interface tools" so that high-risk AI systems can be "effectively overseen by natural persons." Similarly, it also requires that individuals are aware of the possibility of relying and over-relying on the system, and "be able to intervene in the operation."[198]

Because the navigation software is a high-risk system, the conformity assessment would show whether the human interface meets the EU standards, which most likely follow the state of scientific and technical knowledge. If it does, it will most likely be uphill for the plaintiff to prove that the interface is not of the kind the public at large expects and reasonable for the end user. If the conformity assessment is non-compliant with the safety standards, causality will be presumed and the manufacturer or provider will have to prove that this didn't cause the accident (regardless of administrative complaints that may be filed aside, under the AI Act, for nonconformity).

In all cases, if the plaintiff did not abide by her expected standard of care and, for example, didn't follow instructions, was distracted, or was in breach of a legal obligation, the liability of the manufacturer could be reduced, but most likely not eliminated.[199] This is positive, as it would also encourage harm-reducing behavior from AI system end-users.[200] If the plaintiff contributed to the accident with her action or omission with no fault - perhaps she did receive control of the car, but given how control was handed it was not reasonable to expect from her that she would control the vehicle - the Revised PLD also establishes that this should not reduce the liability of the producer.[201]

4.2. Analysis of the example

The example above reveals a few interesting ways in which the Revised PLD complements the AI Act and two important shortcomings:

First, the disclosure of evidence requirement strongly relies on the presumption that extensive evidence will exist. Under the AI Act, however, only the producers and deployers of high-risk systems and foundational models are required to produce and keep documentation about the functioning of AI systems. Recall, that one of the advantages of regulation, according to Shavell, is to mandate the production of information that is not produced.[202] Thus, even if under the PLD courts are empowered to order the defendant to disclose relevant evidence from all AI producers, it is less clear that victims of harm by non-high-risk systems will have access to equivalent evidence than victims of harms by high-risk AI systems. When an accident involves an AI-powered product that falls outside the high-risk system category defined by the AI Act the level of protection may thus be lower, simply because less documentation, and technical standards, may be available. This is a function of the AI Act's structure and less so the liability regime itself.

Second, the ease with which victims will be able to succeed at their liability claims may strongly depend on compliance with the special requirements and standards mandated by the AI Act. This is, again, a residual effect of the AI Act that spills into the liability regime. This was

[183] Proposal New PLD (n136) Article 5; Article 9.1.

[184] Software supporting motor vehicles is under the current high-risk category of the AI Act, but it is also expected that specific regulations will be developed. See Fernandez Llorca and Gomez Gutierrez (n176).

[185] See above the discussion on the AI Act and conformity assessments.

[186] Proposal New PLD (n136) Article 6.1.

[187] Proposal New PLD (n136) Article 6.1(a).

[188] Proposal New PLD (n136) Article 6.1(b).

[189] Proposal New PLD (n136) Article 6.1(f).

[190] Proposal New PLD (n136), Article 6.1(h).

[191] Proposal New PLD (n136) Article 9.1 (b), (c).

[192] Proposal New PLD (n136) Article 10.1(c).

[193] Proposal New PLD (n136) Article 10.1(d).

[194] Proposal New PLD (n136) Article 10.1(e).

[195] See above Part 1.

[196] Proposal New PLD (n136) 243, Article 8.

[197] See AI Act (n3) Article 11.

[198] See AI Act (n3) Article 14.

[199] Proposal New PLD (n136) Art 12.2 this echoes the principle of the contributory conduct or activity of the victim see PETL (n2) Article 8:101.

[200] See Buiten and others (n58).

[201] Proposal AILD (n35) Article 12.

[202] See Part 3.1(b).

exemplified above, and will be important, in the case of hybrid systems under the PLD: When the AI Act is in place, high-risk systems will be very likely to be designed to meet the expectations and standards of the human control requirement. This should improve the interface overall and to a certain degree link the standard of conduct of developers and providers to more clearly defined industry standards. Thus, if the human operator has, for example, not had access to information about the AI system in their training or in the form of readable instructions the AI manufacturer may be held liable.[203] A significant amount of the legal work of proving a defect will thus be focused on proving that the human-AI interface was not fit for purpose. As above, however, if the system at issue is not a high-risk system, less extensive and accurate documentation may be available to prove such claims.

At the same time, recall that in instances where compliance with standards is what led to the harm, developers and deployers will not be held liable.[204] Though this makes sense from the developer and deployers' perspective, it shifts attention to how the human in the loop requirement will be developed in the standard-setting process. If these standardization process fails to account for the difficulties discussed in Part 2, then the outcome will be undesirable and victims are likely to remain unprotected under civil liability rules vis à vis victims of harms that occur without a hybrid AI system: developers will argue that the human was a regulatory requirement, and the human (or their employer) may be able to argue that the system was not fit for purpose.

4.3. Variations on the main theme: AI and safety with a human under the AILD

Now let's assume the situation is similar but the victim is not a natural person or a legal entity. Imagine that the accident involves a semi-automated vehicle operating under human supervision in an industrial setting. The vehicle swerved into a curb, the human operator didn't manage to take control of the vehicle, and this caused an accident which resulted in material damages for the factory-owner. Here, because the victim of harm is a legal entity, the AILD applies.

When we look at how the AILD would perform, it becomes evident that the cliff effects from the AI Act are even stronger on the AILD than on the PLD.[205] The AILD's provision providing for disclosure of documentation only applies to high-risk systems. Victims of harms that occur by or with the participation of an AI system that is not high risk, but is still opaque or complex, will thus still face significant hurdles in overcoming the technical and organizational opacity of AI systems. Additionally, courts and plaintiffs may face a significant challenge of unknown unknowns when trying to order only the "necessary and proportionate" evidence to support a potential claim fault.[206]

In situations where the AILD would apply, there is also the question of the human in the loop. The AILD does not apply "when the damage is caused by a human assessment followed by a human act or omission."[207] The phrasing of the recital does not seem to consider yet the complexities of human assessment after an AI system provides advice. It is unclear how this recital may affect situations where humans and AI are supposed to work together.

Take an illustrative example: In a famous aeroplane crash involving an automated aviation system and a pilot, the accident happened because the pilot failed to steer the plane up, while the system was (wrongfully) steering it down.[208] In such cases of complex interactions, the AILD will apply if the plaintiffs succeed at arguing that this scenario *is not* an instance where "damage is caused by a human assessment followed by a human act or omission, while the AI system only provides information or advice."[209] It may not apply, however, to instances where control is handed over a second before an accident happens, as it often happens on car accidents, and this is considered to comply with best practices and efforts. This is, unless the AILD introduces some of the nuances the newer version of the AI Act has, but plaintiffs will still need to assert and substantiate the likelihood that the human-machine system did not adequately prepare the human for effective control of the situation to establish the applicability of the AILD. In what seems like a circular situation, plaintiffs will only then be able to compel AI developers – or courts - to disclose pertinent evidence. Yet, to be able to assert that, they would benefit from examining the documentation of the human-AI interface and system dynamics.

4.4. Harms to fundamental rights

For a last scenario, let's look at how the system will fare in the case of fundamental rights violations. Recall that one of the main objectives of the AI regulatory framework in general, and the AILD specifically is to help protect and give redress to victims of harm to fundamental rights, such as the right to non-discrimination.[210] It is also worth recalling, however, that EU fundamental rights law is generally applicable mainly to institutions and body of the EU, and to Member States only when they are implementing Union law.[211] This is, of course, unless there are other, specific laws such as data protection law or antidiscrimination law that extend the obligation to private parties to comply with fundamental rights law to private parties.[212] Additionally, Member States have rich traditions on the application of fundamental rights and, in general, it is up to Member States to establish procedural rules for the actions intended to safeguard fundamental rights.[213] The application of the AILD will thus be subject to national, or special, rules on the application of liability law to guarantee the protect fundamental rights.

As with all forms of liability law, victims of fundamental rights in situations that involve an AI system will have to show that harm occurred. This is not necessarily straightforward: Plaintiffs must have legal knowledge or reasonable suspicion of harm and provide sufficient facts and evidence to support the likelihood of a damages claim. However, victims of AI discrimination, for example, may be unaware or

[203] See discussion in Part 1.

[204] Recall that under the PLD, manufacturers and distributors will not be liable if they are able to prove that the defectiveness is due to compliance of the product with mandatory regulations. Proposal New PLD (n136) Article 10(d).

[205] See Hacker (n174) 20, arguing that this problem arises because the EU AI liability regime excessively relies on the risk categories defined in the AI Act and arguing that the list of the AI Act is both over and under inclusive.

[206] See Hacker (n174) 20, making a similar critique and arguing that the list of the AI Act is both over and under inclusive.

[207] Proposal AILD (n159) Recital 15.

[208] Dominic Gates and Lewis Kamb, "Indonesia's devastating final report blames Boeing 737 MAX design, certification in Lion Air Crash" (*The Seattle Time*, Oct. 24, 2019) <https://www.seattletimes.com/business/boeing-aerospace/indonesias-investigation-of-lion-air-737-max-crash-faults-boeing-design-and-faa-certification-as-well-as-airlines-maintenance-and-pilot-errors/ > accessed august 26, 2023.

[209] Proposal AILD (n159) Recital 15.

[210] Proposal AILD (n159).

[211] Charter of Fundamental Rights Art. 51.

[212] See Hacker (n174).

[213] See judgments of 13 December 2017, El Hassani, C-403/16, EU:C:2017:960, paragraph 26, and of 15 September 2022, Uniqa Versicherungen, C-18/21, EU:C:2022:682, paragraph 36.

suspicious of whether an AI system's decision stems from algorithmic bias leading to unlawful discrimination, and they typically cannot access the necessary information from the system's output logs. In some instances, such as the rejection of a loan application, there may be incentives to investigate. Ziosi et al. argue, for example, that discrimination's impact can be more subtle, such as when women consistently receive fewer job opportunities than men. In such cases, discrimination manifests as a lack of opportunity rather than a direct denial.[214]

Additionally, and in some cases, the affectation of a fundamental right may not necessarily amount to damage. Take, for example, the case of data protection law. The GDPR establishes in Article 82(1) that "[a]ny person who has suffered material or non-material damage as a result of an infringement of this Regulation shall have the right to receive compensation from the controller or processor for the damage suffered."[215] The ECJ has explained, however, that the conditions that give rise to compensation for an infringement on an individual's data protection rights require establishing, in essence, similar conditions to any other liability claim: "namely processing of personal data that infringes the provisions of the GDPR, damage suffered by the data subject, and a causal link between that unlawful processing and that damage."[216] One of the reason for this is that the GDPR, specifically, provides for administrative and judicial remedies before a supervisory authority in case of an infringement of the GDPR some of which have a punitive purpose and are not conditioned by the existence of damage.[217] Thus, to be able to obtain compensation, the injured party must prove that the consequence of the breach of the GDPR constituted a certain form of damage, even if a non-material damage (which the court has also explained must be interpreted broadly).[218]

As in the previous example, victims seeking compensation for infringements on their fundamental rights may not always be able to do so under the design of the legal system, unless there is additional harm. Even when they can, they may encounter challenges in accessing and understanding relevant evidence: Under the AILD, plaintiffs have a right to access evidence about high-risk AI systems which they suspect caused them harm. This right requires that plaintiffs present "facts and evidence sufficient to support the plausibility of a claim,"[219] is limited to evidence that is "necessary and proportionate" to support the claim,[220] and requires courts to only order the disclosure of evidence when claimants have made "all proportionate attempts at gathering the relevant evidence from the defendant."[221] As Ziosi and co-authors explain, it may be hard for non-experts to consider what evidence can be considered plausible, and presume victims awareness of harm.[222] Additionally, claimants may face difficulties proving fault not only because of the legal nature of documentation, when they exist, may still be challenging for less technically literate plaintiffs.[223]

Lastly, and as already discussed above, however, the AILD is not supposed to apply to situations caused by a human assessment followed by a human act or omission where the AI system only provides information or advice.[224] Though high-risk systems are not the only type of AI system that can eventually affect fundamental rights, to the extent the list of high-risk systems contains a list of the "usual suspects," it seems like a notable exclusion.[225] This is paradoxical as a central objective of the whole AI regime in Europe is to protect and mitigate fundamental rights related harms.[226]

4.5. Conclusion to this part

This Part "ran" the EU liability directives through three examples of situations where AI harms occurred, and a human was involved: two safety harms and harm to fundamental rights.

Based on these examples and applying the framework laid out in Part 2, the following conclusions can be drawn:

First, the liability regime seems to successfully complement the AI Act, especially in instances where high-risk systems are involved. This occurs for two main reasons: because most of the rules directed at facilitating access to evidence are directed at high-risk systems, and because under the AI Act, it is only developers of high-risk systems that will produce the desirable information. In some instances, it may be unrealistic to expect victims of harm involving high-risk systems to assess the technical documentation and prove, for example, lack of comformity. An additional downside of this complementarity is that victims of harm by AI systems that are complex, opaque, or autonomous and not defined as high-risk systems or foundation models under the AI Act, may still face important obstacles when trying to prove the existence of a defect (in the case of "products," or other special regimes), or fault on the side of the producer or deployer (in all other cases). The AI Act tiered framework thus drifts over the AI liability regime and the trustworthy AI regime, and most of the incentives to take care and produce desirable information fall up to the developers and deployers of high-risk systems.

Second, the liability regimes treat human-hybrid systems in a contradictory manner. On the one hand, the AI Act mandates human oversight over high-risk systems and emphasizes how the human-AI interface must be designed to be effective. Increased focus on whether a human-AI interface is "fit for purpose" is an important improvement from the status quo. However, human-AI interactions are complex and not always desirable, and the focus of the AI Act on human control may lead to situations where, under the AILD, designers may rightfully claim that the defect or situation at issue arose *because* they must comply with the human supervision requirement.

On the other hand, the AILD excludes from its coverage systems where AI is advisory, rather than decisionmaker and thus developers and deployers of high-risk systems may not be subject to the AILD at all. Paradoxically, this will be the case for many systems used to make

[214] Ziosi (n71).

[215] Regulation (EU) 2016/ 679 General Data Protection Regulation, OJ L 119, 4.5.2016, Article 82 (GDPR).

[216] Case C-300/21, *UI v. Österreichische Post AG* [2023] ECLI:EU:C:2023:370, 14 (UI v. OP) 36.

[217] UI v. OP (n216) 40, GDPR (n227) Article 83 and 84.

[218] UI v. OP (n216) 50 (in the decision it is unclear what is a damage within the meaning of the GDPR).

[219] AILD, Article 3(1).

[220] AILD, Article 3(4).

[221] AILD, Article 3(2).

[222] Ziosi (n71) 7.

[223] Ziosi (n71)7.

[224] Proposal AILD (n159) Recital 15.

[225] High risk systems include AI systems used in critical infrastructures; educational or vocational training; safety components of products; employment, management of workers and access to self-employment (e.g. CV-sorting software for recruitment procedures); essential private and public services; law enforcement that may interfere with people's fundamental rights; migration, asylum and border control management; administration of justice and democratic processes. See AI Act (n3) Annex III.

[226] See discussion in Part 2.

decisions that are consequential to fundamental rights, such as systems used in educational and vocational settings to determine who can access a certain program.

Third, the AILD may be a somewhat limited measure to solve the individual redress and recourse gap in the case of affectations to fundamental rights, even when they apply. Victims may, however, face difficulties identifying that they suffered harm to their fundamental rights and even when they do the application of the AILD will be constrained to Member State or specific regulation regarding the use of liability law to seek redress for harms to fundamental rights. EU Law and Member State liability laws typically distinguish between the affectation of a fundamental right and the effective occurrence of material or immaterial harm to grant compensation, which is both hard to prove and may not always occur.

Lastly, the effectiveness of the liability regime seems to importantly rely on the development of standards that are mandated under the AI Act. This is true for the human oversight requirement, as I showed above, but it may be true for most other requirements where the development of standards will lead to a better understanding what "best practices" around the development and deployment of AI systems should be. Indeed, these standards could be used to establish duties of care. This also highlights the importance of standards to the overall effectiveness, and democratic legitimacy, of AI governance.[227]

Based on these observations, the next and last section offers some recommendations for addressing these limitations in the current AI Liability framework in Europe.

5. Suggested reforms and key elements for the broader discussion

The proposed revised PLD and the AILD seek to update the existing liability frameworks in EU Member States so that individuals who suffer such harm obtain fair compensation, and thus ensure, in general, that the uptake of AI is done with individual interests in mind. As the EU strategy emphasizes, and to the extent the EU also wants to incentivize the development and adoption of "trustworthy" AI, a fit-for-purpose liability regime also creates legal certainty for businesses.[228]

The proposals, though certainly advancing in an important direction and part of a broader regulatory initiative. This Part proposes a few avenues in which the AI Liability Regime can be further improved, based on the considerations of the previous parts, to (1) better address the information asymmetries for systems that are *not* subject to special requirements under the AI Act; (2) ensure victims of harms in AI-Human systems are not left worse off than victims of solely automatized or non-automatized systems; (3) improve the redress of fundamental rights and create better incentives for AI developers and deployers to exercise more care.

5.1. Addressing information asymmetries

Information asymmetries between plaintiffs and AI developers and producers are a function of AI opacity because it obstructs effective inspection of AI systems.[229] The EU proposals successfully address organizational opacity, especially for high-risk systems under the AILD and, in general, under the PLD, because developers will no longer be able to assert confidentiality over the evidence. Thus, once the AI Act is applicable, there will also be adequate documentation, which should also diminish the difficulty in scrutinizing the workings of a system. Similarly, the strict liability regime under the PLD, and the rebuttable presumption of causality under the AI Act, are positive adjustments to ease the burden of victims of proving causality.

However, it is noteworthy that the proposed regime may better serve the victims' high-risk systems and foundation models, as defined by the AI Act, than the victims of harm by other systems. From an organizational opacity perspective, in the case of the AILD, the courts' power to demand the disclosure of relevant evidence extends only to high-risk systems. Even though the AI Act's high-risk systems list is a good proxy for the systems that are most likely to cause harm, and are complex, they will not be the only systems that cause harm, nor are they the only opaque and complex systems that may, both now and in the future, cause harm. Thus, it is advisable that under the AILD, as in the PLD, courts are *always* empowered to order the defendant to disclose relevant evidence that is at its disposal, upon request of an injured person claiming compensation and when the claimant has presented facts and evidence sufficient to support the plausibility of the claim for compensation.

In the case of technically opaque or complex systems, victims seeking to prove fault under the AILD may again find it easier when the system is high-risk this is considering that explanatory documentation that can be relied upon to provide evidence will most likely be the one produced on the transparency, explainability and record-keeping requirements produced under the AI Act for high-risk systems. Additionally, the development of legal and industry standards will enable plaintiffs to compare a producer or deployer's behavior with other actor's behaviors and standards of care.

Consequently, if the power to request documentation from non-high-risk systems is extended, courts could request developers and deployers to provide ex-post explanations of how a system operates. This should be done to the extent possible and based on a reasonable justification presented by the plaintiff as to why this is needed.

5.2. Human-AI hybrid systems and the role of standards

In instances where liability claims involve human-AI hybrid systems, courts should emphasize evaluating the identity of the human-AI interface. This is particularly crucial when examining cases where the human element in the loop is being considered as the cause or a contributing factor to AI-related harm.

To shift legal processes in this direction, and as the European Union's framework for trustworthy AI reaches completion, these considerations must be considered during the process of establishing industry standards for the human supervision requirement under the AI Act. Indeed, the standard-setting process will play a structural role not only in implementing and materializing the ambitions of the AI Act but, importantly, in creating the baseline expectations to assess and evaluate liability claims.

Human oversight standards must, for example, mandate a clear definition of the roles and responsibilities of each party involved, consider the level of training and automation of the system in place, and account for the competencies possessed by the human actor in question. Similarly, standard setting bodies should mandate, for example, that depending on the competences of the expected AI users, and the sensitivity of the situation, trainings and clear instructions are part of

[227] On the importance of standards for the implementation of the AI Act see Edwards (n6); Michael Veale and Zuiderveen Borgesius, "Demystifying the Draft EU Artificial Intelligence Act - Analyzing the good, the bad, and the unclear elements of the proposed approach," (2021) 22(4) CLRI,t 8, 9; Mélanie Gornet and Winston Maxwell, "The European approach to regulating AI through technical standards" (On file with the authors, 2024).

[228] White Paper on AI (n8) 13.

[229] See making a similar argument Commission Report on safety and liability implications of AI (n21) 16.

teaching humans how to operate AI systems. Professionals such as pilots or machine operators should arguably be held to a more stringent accountability standard compared to everyday consumers.

As in critical safety industries or other industries with experience in human-machine interactions, EU standard-setting bodies and judges should pay special attention to the stated goals of the AI-Human system, the reasonability of those expectations, and systems are designed and labelled sufficiently for effective use, and address training and organizational policies.[230] Though from a liability perspective technical standards are different from standards of care, it seems inescapable that at least part of the evaluation of compliance with standards of care will rely on what are defined to be the appropriate technical standards for hybrid systems.

5.3. Redress to affectations of fundamental rights

The third element of discussion is the suitability of the AILD to seek redress for affectations of fundamental rights.

The first shortcoming of the existing AILD is the exclusion from its scope of application of AI systems are supervised by humans. This would lead, for example, to an algorithm like the one at issue in the Dutch scandal outside of its scope of application, but it is also especially worrisome as human supervisors are increasingly introduced to specifically mitigate the risks posed by AI systems used in different forms of decision-making that can affect fundamental rights. A first key recommendation is, thus, to eliminate this requirement.

The second, more structural shortcoming, is that liability law necessarily requires the occurrence of harms to warrant compensation - the main remedy within liability law. This may constrain the AILD's capacity to facilitate victims' access to justice and may create fewer incentives for certain AI providers to take optimal care.

To be fair, the general framework for the trustworthy AI Act is centred around the understanding that the protection of fundamental rights isn't only about an individual's right itself - for example, a person's right to equality before the law -, but it also about building societies that are respectful of fundamental rights. The EU's system of fundamental rights seeks to achieve this, by, for example, promoting political participation and a functioning democracy and directing the work of different government bodies towards building societies and markets where fundamental rights are in general guaranteed. These systemic aspects of the protection of fundamental rights are the objectives to be addressed via the enforcement of the AI Act and its safety requirements. Additionally, in some instances fundamental rights violations are better addressed with non-pecuniary remedies, such as injunctions, declarations, or specific performance orders to correct the violation.[231]

At the same time, one of the key concerns of civil society is the lack of mechanisms of robust mechanisms for redress for individuals and groups affected by AI systems.[232] Even if the final version of the AI Act includes a remedy chapter that includes a right to lodge complaints with a market surveillance authority, it remains unclear what the effectiveness of this mechanism will be and how it will act as a mechanism to compensate for individual affectations to fundamental rights.[233] To improve access to recourse for individuals who are victims of illegal violations of fundamental rights in situations involving AI systems, European and Member State authorities may consider adopting or expanding mechanisms like those of the AILD within other procedures intended to effectively address such violations

6. Conclusion

While AI can do much good, it can also harm. The characteristics of AI, and how individuals participate in and interact with AI systems make it difficult to trace back potentially problematic decisions or outcomes made with their involvement. This makes it difficult for victims of harm to obtain redress. The 2022 directives proposed by the European Commission seek to update the existing liability frameworks in EU Member States so that victims of harm with an AI system obtain fair compensation, and thus to ensure, in general, that the uptake of AI is done with individual interests in mind and with legal certainty for businesses.[234]

The proposals are an important complement to the AI Act's risk and safety approach. Indeed, relying solely on risk regulation has distributive consequences, including the possibility that individual harms and costs will be dismissed if a particular measure makes sense collectively, which may especially harm minorities.[235] It may also lead to situations where, because regulators are fallible, organizations don't have enough incentives to take optimal care.[236] Similarly, one of the main arguments that were raised when the AI Act was first published was that it didn't include individual rights nor rights of action for affected persons, even if its stated goal is to protect fundamental rights in Europe.[237] In this context, liability law becomes an important vehicle to ensure that the vast and fast adoption of AI systems in all facets of life and society is done in a way that guarantees the protection of people's rights and interests, but also to provide legal certainty for AI developers and deployers. It is, also, an important moment of policy choice, where not only the interests of victims but also the societal interests in adopting and developing AI are weighed against each other.

Nevertheless, this Article has shown that the AILD and the PLD, in their current forms, fall somewhat short of their ambition to effectively complement the AI Act, not in small part because they very strongly rely on the tiered framework developed by the AI Act. This occurs, especially, because the *ex-ante* regulatory interventions will often lead to the creation of the documentation, standards and information that will be important to successfully succeed in liability claims *ex post*. This is especially the case for hybrid systems. This analysis also calls into question whether liability law is the best mechanism to give victims of affectations to their fundamental rights, when an AI system is involved, a viable mechanism to seek redress.

The time is right, however, for the EU Commission and Parliament, and legislators around the world, to have a broader conversation about the scope of liability and individual redress mechanisms for AI-related harms. In the EU, some of the elements identified here may be an unavoidable result of focusing attention on a particular and limited set of systems in the AI Act. Other issues – such as the standard-setting process – are outside the scope of the specific conversation on liability but will be critical to its successful implementation. EU institutions, however, should extend some of the benefits proposed by the AILD and the PLD to more or all harms involving opaque and complex AI systems, extend the application of the AILD for all AI systems regardless of whether a human is supervising, and explore other avenues for individuals to seek redress for AI affectations to their fundamental rights. Doing so will better enable the goal of trustworthy AI and help realize the EU Approach to Artificial Intelligence, which focuses on fostering trust, enhancing

[230] Crootof and others (n27) 466.

[231] UI v. OP (n223) 39, GDPR (n227) Article 77.

[232] EDRi and coalition partners (n131).

[233] EDRi and coalition partners (n131).

[234] White Paper on AI (n8) 13.

[235] Kaminski (n4), 8.

[236] See discussion of law and economics analysis of regulation in Part 2.

[237] EDRi and coalition partners (n131).

research and industrial capacity, and ensuring safety and fundamental rights. Similarly, as other countries pass AI regulations, the example of the EU liability framework for AI may be useful to analyze to better understand how liability law can complement AI risk regulations.

Declaration of competing interest

The author declares 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.

Acknowledgments

Thank you to Margot Kaminski, Maximilian Gahntz, and the participants of WeRobot 2023 for their comments and feedback on previous versions of these piece. Thank you, also, to the editors and reviewers of CLSR, and to Francesca Elli and Giovanna Hajdu Hungria Da Custódia for their research assistance.