International Journal of Law and Information Technology, 2024, **32**, eaae021
https://doi.org/10.1093/ijlit/eaae021
Advance access publication 14 September 2024
Article

Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive

Guido Noto La Diega[*], Leonardo C.T. Bezerra[†]

ABSTRACT

In Europe, the governance discourse surrounding artificial intelligence (AI) has been predominantly centred on the AI Act, with a proliferation of books, certification courses, and discussions emerging even before its adoption. This narrow focus has overshadowed other crucial regulatory interventions that promise to fundamentally shape AI. This article highlights the proposed EU AI liability directive (AILD), the first attempt to harmonize general tort law in response to AI-related threats, addressing critical issues such as evidence discovery and causal links. As AI risks proliferate, this article argues for the necessity of a responsive system to adequately address AI harms as they arise. AI safety and responsible AI, central themes in current regulatory discussions, must be prioritized, with ex-post liability in tort playing a crucial role in achieving these objectives. This is particularly pertinent as AI systems become more autonomous and unpredictable, rendering the ex-ante risk assessments mandated by the AI Act insufficient. The AILD's focus on fault and its limited scope is also inadequate. The proposed easing of the burden of proof for victims of AI, through enhanced discovery rules and presumptions of causal links, is insufficient in a context where Large Language Models exhibit unpredictable behaviours and humans increasingly rely on autonomous agents for complex tasks. Moreover, the AILD's reliance on the concept of risk, inherited from the AI Act, is misplaced, as tort liability intervenes after the risk has materialized. However, the inherent risks in AI systems could justify EU harmonization of AI torts in the direction of strict liability. Bridging the liability gap will enhance AI safety and responsibility, better protect individuals from AI harms, and ensure that tort law remains a vital regulatory tool.

KEYWORDS: AI; generative AI (GenAI); liability; torts; harmonization; AI liability directive.

[*] Guido Noto La Diega, Professor of Law, Technology and Innovation at the University of Strathclyde, Glasgow, where they lead the LLM/MSc Law, Technology and Innovation, and the namesake research theme. School of Law, University of Strathclyde, The Lord Hope Building, 141 St James Rd, Glasgow G4 0LT, United Kingdom. Tel: +441414448427. Email: guido.notoladiega@strath.ac.uk.

[†] Leonardo Teonacio Bezerra, Lecturer in A.I/Data Science at the University of Stirling. Division of Computing Science and Mathematics, University of Stirling, Cottrell Building, Stirling FK9 4LA, United Kingdom. Tel: +44176467421. Email: leonardo.bezerra@stir.ac.uk. This is a genuinely collaborative work, Noto La Diega is responsible for Sections 1, 3, 4, 5 and Bezerra for Section 2.

"Our machines are disturbingly lively, and we ourselves frighteningly inert"
Donna Haraway, *Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s* (1985)

INTRODUCTION

For many years, the artificial intelligence (AI) governance discourse has been dominated by a diffidence towards are regulation and a preference for non-binding initiatives such as ethical charters, and manifestos. In an unexpected turn of events, many countries are now embracing the idea of AI-related hard laws.[1] In this context, AI safety and responsible AI appear to be the commanding concepts.[2] As we write, the European Parliament has approved the AI Act,[3] which is often described as the application of the product safety model to AI[4], and heralded as the 'world's first binding law on artificial intelligence [...] a new model of governance built around technology'.[5] Leaving aside the veracity of the first statement,[6] and the positivity of the latter,[7] we have noted elsewhere that one of the main critical aspects of the AI Act is it was built around narrow AI, with a clumsy last-minute attempt of shoe-horning generative AI (GenAI) into the new law, rather than rewriting it from scratch.[8] The AI Act has also been criticised because it focuses on the impossible task of predicting the impact of the AI system on fundamental rights, thus instantiating an ex-ante approach that can be easily bypassed by AI companies, eg by contractually shift liability.[9] At the same time,

[1] Even the US and the UK—whose neoliberal approach has always meant a preference for deregulation and self-regulation—have recently, albeit timidly, embraced a more top-down approach. In October 2023, the President of the USA issued the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. In February 2024, the UK backtracked on its initial plans to leave it to AI businesses to self-regulated, and set up to introduce binding requirements for developers of highly capable general-purpose AI models to ensure their safety (Department for Science, Innovation & Technology, 'A Pro-Innovation Approach to AI Regulation—Government Response to Consultation' (2024) CP 1019, 4).

[2] For example, it seems clear that the UK approach to AI governance is dominated by the pursuit of safety (see eg the aforementioned legislative initiatives as well as the AI Safety Summit). One could speculate that, whereas ethics and fundamental rights have become increasingly contentious, safety is a ... safer notion, as one could hardly imagine any argument in favour of unsafe AI. We should be careful, however, as the single-minded focus on safety may lead to overlooking the wider societal risks associated to this technology.

[3] As we write, the Regulation of the European Parliament and of the Council on laying down harmonized rules on Artificial Intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139, and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797, and (EU) 2020/1828 (Artificial Intelligence Act) has not been published in the Official Journal. The final text is however available at <https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/en/pdf> accessed 17 June 2024.

[4] Tycho de Graaf and Gitta Veldt, 'The AI Act and Its Impact on Product Safety, Contracts and Liability' (2022) 30 Eur Rev Private Law 803. This will be a *lex specialis* in relation to the new Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council and Directive (EU) 2020/1828 of the European Parliament and the Council, and repealing Directive 2001/95/EC of the European Parliament and of the Council and Council Directive 87/357/EEC [2023] OJ L 135/1.

[5] 'Artificial Intelligence Act: MEPs Adopt Landmark Law' (*European Parliament*, 13 March 2024) <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law> accessed 15 March 2024.

[6] Other binding laws already existed. Notably, China adopted a number of regulations, including Generative AI Regulation, which came into force on 15th August 2023 (see Jing Cheng and Jinghan Zeng, 'Shaping AI's Future? China in Global AI Governance' (2023) 32 Journal of Contemporary China 794.

[7] For many years, technological neutrality has been one of the constitutional principles in the field of Internet governance. While not immune to criticism (eg Chris Reed, 'Taking Sides on Technology Neutrality' (2007) 4 SCRIPTed 263), this approach has so far enabled the law to retain the flexibility required to remain relevant despite the pace of technology development (see Joshua AT Fairfield, *Runaway Technology: Can Law Keep Up?* (Cambridge University Press, Cambridge (UK) 2021). It remains to be seen if the same can be said of this new generation of technology-specific law.

[8] Guido Noto La Diega and Christof Koolen, 'Generative AI, Education, and Copyright Law: An Empirical Study of Policymaking in UK Universities' 2024 EIPR.

[9] Martin Kretschmer and others, 'The Risks of Risk-Based AI Regulation: Taking Liability Seriously' (2023) DP8517 CEPR Discussion Paper No. 18517. CEPR Press, Paris & London 10 <https://cepr.org/publications/dp18517> accessed 15 March 2024.

the AI Act can be seen as a codification of—and can be complemented and operationalized by—'responsible AI'[10] initiatives, ie bottom-up initiatives normally backed by national or transnational governance bodies aimed at 'providing concrete recommendations, standards and policy suggestions to support the development, deployment and use of AI systems'[11]. Responsible AI requires robust forward-looking governance, and at its core there must be questions of who should be liable if AI harms humans and under which circumstances.[12] We posit that there can be no responsible AI without AI liability. There can also be no AI safety without AI liability, ie a clear and comprehensive liability framework for AI, one that would provide strong incentives to develop and deploy systems that are safe by giving victims easy ways to access compensation.

Considering the limitations of the AI Act and the importance of liability for responsible AI, a framework of ex-post liability rules that intervenes if something does go wrong has been said to be flexible enough to accommodate future AI developments.[13] We concur. AI harms—from disinformation to manipulation through to discrimination—are multiplying, to the point that, in November 2023, the OECD launched its AI Incidents Monitor.[14] This tool has recorded over 11,000 incidents and hazards so far, which explains why compensating the harms that occur should be a primary concern if responsible AI and AI safety are to become a reality. As calls to develop responsible AI multiply, there is a lack of standardization in responsible AI reporting, with OpenAI, Google, Anthropic, and the other leading developers primarily testing their models against inconsistent responsible AI benchmarks, which can in turn lead to range of vulnerabilities and flaws.[15] To develop and deploy responsible AI systems requires dealing with the question of how we deal with AI-generated harms; this is pressing as the pace of the evolution in this space means that 'safety culture and processes have taken a backseat to shiny products'[16]. To incentivize the responsible development and deployment of safe AI systems, the harmonization of tort law heralded by the draft AI Liability Directive (hereinafter AILD)[17] can play a vital role. Indeed, tort liability—also known as civil, non-contractual, extra-contractual, delictual, and Aquilian liability[18]—performs the two-fold function of providing tortfeasors (eg the developer of a facial recognition system) with 'an incentive to optimize their level of care and level of activity, but it can also serve as an important tool to compensate injured persons'[19] (eg the black migrant that is targeted by the police due to algorithmic bias in the AI system). We put forward that bridging the liability gap will help achieve AI safety and responsible AI, which in

[10] To a large extent, this phrase corresponds to the idea of trustworthy AI, which is more common in the EU. Among the reasons to prefer the former to the latter, 'trustworthy AI' is usually linked to ethics, which is a contentious area, and 'trustworthiness' is a slippery concept as it refers to a particular kind of behaviour that is considered to be good when it is displayed by individuals or organizations. See Charlotte Stix, 'Artificial Intelligence by Any Other Name: A Brief History of the Conceptualization of "Trustworthy Artificial Intelligence"' (2022) 2 Disc Artif Intell 26, para 2.4.

[11] Virginia Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way* (Springer, Cham (Switzerland) 2019) 95. The author mentions some examples, IEEE initiative on Ethics of Autonomous and Intelligent Systems and the ethical guidelines of the Japanese Society for Artificial Intelligence.

[12] ibid 104.

[13] Kretschmer and others (n 12) 10.

[14] 'OECD AI Incidents Monitor (AIM)' (*OECD*) <https://oecd.ai/en/incidents> accessed 18 June 2024.

[15] HAI, 'Artificial Intelligence Index Report 2024' (2024) Stanford Univ Human-Centered Artif Intell 17 <https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf>.

[16] A former senior employee of OpenAI cited in Dan Milmo, 'OpenAI Putting "Shiny Products" above Safety, Says Departing Researcher' *The Observer* (18 May 2024) <https://www.theguardian.com/technology/article/2024/may/18/openai-putting-shiny-products-above-safety-says-departing-researcher> accessed 23 May 2024.

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

[18] These are not perfectly overlapping synonyms, but they tend to have in common the reference to the private law rules that can be invoked when a harm occurs, with the exception of contractual issues (eg liability for breach of contract). See eg, Percy H Winfield, *The Province of the Law of Tort* (Cambridge University Press, Cambridge (UK); The Macmillan Company, New York City (NY) 1931) 32.

[19] Shu Li and Béatrice Schütte, 'The Proposed EU Artificial Intelligence Liability Directive: Does/Will Its Content Reflect Its Ambition?' [2024] Technol Regul 143.

turn will serve the two-fold goal of better protecting people from AI and helping tort law retain its relevance as a key regulatory tool.

Against this backdrop, this article—adopting a doctrinal method that focuses on European private law, with insight from Italy and the UK—pursues a two-fold objective. First, it critically assesses whether AI responsibility and safety can be achieved by the proposed AILD which is tasked with bridging the AI liability gap through enhanced discovery rules and a presumption of causal link between fault and AI-generated damage[20]. Second, as the proposal was presented two months before ChatGPT became commercially available,[21] this paper scrutinizes whether the AILD is fit for GenAI and its harms. Indeed, the draft AILD is informed by the EU's *White Paper on Artificial Intelligence*, which was focussed on the risks associated with certain specific uses and applications of AI,[22] rather than on general purpose AI models. To achieve these objectives, the next section will briefly introduce the technological underpinnings of GenAI and related harms. We will then analyse the proposed AILD framework to evaluate whether it can be pivotal to responsible AI by incentivizing its safety through ex-post liability rules. Finally, we will check the fitness of the draft AILD vis-à-vis GenAI, which is pressing as we have time to rewrite the law and avoid its premature obsolescence.

TECHNOLOGICAL UNDERPINNINGS AND POTENTIAL HARMS OF GENAI

AI is a broad field whose origins in the mid-20th century coincided with the meeting between theoretical computer science and the first modern computing machinery[23]. The pivotal role played by AI in World War II led to a subsequent rapid development of the field, with successful applications observed in the industry throughout the second half of the 20th century. However, the progress of AI as a field has never been homogeneous nor linear, with several AI technologies promising breakthroughs that later led to the so-called *AI winters*[24]. For instance, while heuristic optimization in operations research can be regarded as a successful AI approach that keeps automating our industrial and societal processes, approaches such as rule-based expert systems did not become mainstream as originally anticipated. In general, AI sees its best results when (i) its scope is narrow, ie when employed to address very specific tasks, and; (ii) the limitations of the techniques employed are well-understood, which is often not the case given the hype that novel AI breakthroughs stir.

Perhaps the AI field with most breakthroughs and setbacks is machine learning (ML),[25] particularly concerning a family of algorithms called artificial neural networks (ANNs). In detail, ML algorithms are techniques that either try to (i) identify/learn patterns from data or (ii) devise a given behaviour based on interactions with the environment. The earliest implementation of ANNs was deemed revolutionary in the 1960s for its promise to mimic human cognition,[26,27]

[20] For some recommendations on how to change the AILD to better address the liability gap and the information gap see Marta Ziosi and others, 'The EU AI Liability Directive (AILD): Bridging Information Gaps' (2023) 14 Eur J Law Technol 8–9 <https://ejlt.org/index.php/ejlt/article/view/962> accessed 21 June 2024.

[21] The European Commission released the draft AILD on 28th September 2022, ChatGPT was launched on 30th November 2022.

[22] European Commission, 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust' (2020) COM(2020) 65 final 13.

[23] Turing, Alan, 'Computing machinery and intelligence', Mind, 1950.

[24] The term AI *winter* has been coined to describe the periods of retraction in investment on and adoption of AI technologies that followed periods of major interest in the field due to a given breakthrough.

[25] Christopher M. Bishop, *Pattern recognition and machine learning* (Springer, New York (NYC) 2006).

[26] McCulloch, W and Pitts, W. 'A Logical Calculus of Ideas Immanent in Nervous Activity' (1943) 5(4) Bull Math Biophys115–133.

[27] Rosenblatt, F 'The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain' (1958) Psychol Rev 65

only to see the then hype frustrated as theoretical computer science demonstrated its limitations.[28] In the 1980s, the devise of the currently most employed ANN training approach led to another surge in interest,[29] with many specialized ANN architectures[30] that are used to this date being proposed in that period or shortly after. Already in the early 2000s, ML algorithms employed for predictive purposes achieved accurate results when the data used was tabular and of good quality, and over the past decade this also became true in fields where data is not tabular, eg image, text, audio, video, and/or code. Among the core ideas that enabled recent results are (i) big data,[31] the technology to store and process the vast, diverse, and fast-growing data produced in the era of social media, and (ii) deep learning,[32] the technology that employs highly parallel[33] and computation-intensive ANNs. Still, the accuracy observed for those models heavily rely on the concept of narrow AI, as previously discussed. From a responsible AI standpoint, though, big data complicates the *many-hands problem*,[34] as the supply chain surrounding the AI lifecycle becomes considerably long and intricate.[35] In addition, deep learning further increases the already poor transparency of ANNs, leading to an *opacity* that conceals for all practical purposes the reasoning behind the algorithmic decision-making.

GenAI is the most recent ANN breakthrough that promises a revolution and may deliver yet another winter. Broadly, GenAI is a term used to describe ANN algorithms that create content based on the training data used in the development of the generative models. Among the most prominent examples of GenAI technologies are generative adversarial networks (GANs),[36] used for computer vision (images), and large language models (LLMs),[37] used originally for natural language (text). GANs comprise two ANNs that play the adversarial roles of an art forger and an art investigator. A GAN model is regarded as accurate when the investigator can no longer differentiate between original (training) and forged (generated) content. To an extent, GANs still represent narrow AI as they are generally employed to produce images that are similar to their training dataset. In turn, LLMs are obtained from an ANN trained to predict the next token (a word part or a symbol) in a sequence, based on the probability patterns learned from given corpora. Though simple in concept, language models address complex tasks such as question-answering and text summarization, though only recently the same language model started being used for multiple natural language tasks rather than having specific models for specific tasks.

[28] Marvin Minsky and Seymour A. Papert. *Perceptrons: An Introduction to Computational Geometry* (MIT Press, Cambridge (MA) 1969) 480, 479, 104.

[29] David E Rumelhart, Geoffrey E Hinton and Ronald J Williams (1986). 'Learning Representations by Back-Propagating Errors' (6088) 323 Nature 533–536.

[30] Being networks, ANNs can vary as to their topology and the type of computation performed at nodes and/or layers. The resulting architecture of the network is decisive for its performance and varies as a function of the task and data. For images, for instance, convolutional architectures proposed in the 80's gained significant popularity in the 2010's. A similar pattern is observed for recurrent networks, successfully applied to sequential data such as text and audio until recently superseeded by novel architectures such as the Transformer.

[31] Ghemawat, Sanjay, Howard Gobioff and Shun-Tak Leung. 'The Google File System'. In Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles, 2003,pp. 29–43..

[32] Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning* (The MIT Press, Cambridge (MA) 2016).

[33] Parallel computing refers to the ability of executing different parts of an algorithm at the same time, typically through multiple processing units. Moving from sequential to parallel execution of algorithms is non-trivial in several scenarios, and for sequential data such as text the algorithms employed until recently were majorly sequential due to the need to respect the order of the training data.

[34] This goes at the core of the AI accountability gap and it is problematic because '[i]ndividual citizens may have a hard time finding out who they should turn to, if data are incorrect, corrupted, or biased as a collective outcome of a series of minor contributions' (Filippo Santoni de Sio and Giulio Mecacci, 'Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them' (2021) 34 Philos Technol 1057, 1066.).

[35] Crawford, Kate. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press, New Haven (CT) 2021).

[36] Ian Goodfellow and others, 'Generative Adversarial Nets', in *Z. Ghahramani and M. Welling and C. Cortes and N. Lawrence and K.Q. Weinberger* (eds), *Advances in Neural Information Processing Systems* 27 (MIT Press, Cambridge (MA) 2014) <https://papers.nips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html> accessed 25 June 2024.

[37] Shervin Minaee and others, 'Large Language Models: A Survey' (arXiv, 20 February 2024) <http://arxiv.org/abs/2402.06196> accessed 28 June 2024.

Indeed, being trained on ever-larger training datasets, LLMs display capabilities that had not been anticipated (a phenomenon sometimes referred to as *emergence*). Among the most relevant are (i) zero-shot learning, where the model is able to perform a task it has not been trained for, and; (ii) few-shot learning, where the model can be taught to perform a new task through novel examples rather than training. Importantly, the good results observed for text led developers to consider multimodal language models, ie models that can also address image, audio, video, and/or code.

The unexpected capabilities of LLMs have stirred a novel AI rush towards autonomous agents. Historically, the term agents has been used in AI scholarship to refer to systems that are able to make decisions autonomously, whereas the term models refer to the algorithms behind the decision-making. In the current race, agents range from applications that are more connected to natural language processing, eg conversational agents such as ChatGPT, to applications that present an added layer of complexity, eg automating healthcare and law or connecting to external systems such as web search. Both examples further reinforce the scope and many-hand problem issues previously discussed, especially given the expected massive investments for developing and adopting these systems over the next years.[38]

From the high-level description of the technological underpinnings of GenAI given above, one can identify some potential harms they add to the already existing harms discussed for ML algorithms in general. Regarding GANs, the pursuit of models that are able to deceive investigators means that media content they produce is hard to distinguish from media content that has been created by human beings. The public availability of GANs has led to a concerning deepfake industry that is now under public scrutiny due to its potential applications for pornography[39] and identity theft.[40] Regarding multimodal LLMs, having a single model for a diverse range of tasks increases the uncertainty regarding content. Hallucinations, for instance, are content that an LLM creates with no factual source to ground them. Worse, LLMs may create non-existent references, or credit existent references with statements that are in direct opposition to their actual claims.[41]

Adding to the potential harms inherent to the technology behind GenAI are the harms involving model training and deployment. Regarding training, machine learning is highly dependent on the characteristics of the data that is used as input. Currently, AI developers have been collecting vast datasets by scraping internet websites. This has led to significant disputes regarding copyright[42] and privacy,[43] and may incur in other relevant issues such as reproducing existing societal biases and therefore discrimination.[44] As a result, AI companies are trying to

[38] Keach Hagey and Asa Fitch, 'Sam Altman Seeks Trillions of Dollars to Reshape Business of Chips and AI', *Wall Street Journal*, https://www.wsj.com/tech/ai/sam-altman-seeks-trillions-of-dollars-to-reshape-business-of-chips-and-ai-89ab3db0, accessed on 26 June 2024.

[39] 'Government Cracks down on "Deepfakes" Creation' (*GOV.UK*) <https://www.gov.uk/government/news/government-cracks-down-on-deepfakes-creation> accessed 25 June 2024.

[40] US Department of Homeland Security, 'Increasing Threats of Deepfake Identities' <https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf>.

[41] Kevin Wu and others, 'How Well Do LLMs Cite Relevant Medical References? An Evaluation Framework and Analyses' (arXiv, 2 February 2024) <http://arxiv.org/abs/2402.02008> accessed 25 June 2024.

[42] See case law cited in Guido Noto La Diega and Christof Koolen, 'Generative AI, Education, and Copyright Law: An Empirical Study on Policymaking in UK Universities' (2024) 6(6) EIPR 346, most famously *The New York Times Company v Microsoft Corporation*, 1:23-cv-11195, (S.D.N.Y.).

[43] Most recently, the Dutch data protection authority has concluded that most data scraping would potentially be in breach of the EU General Data Protection Regulation (GDPR), as the only available lawful basis for processing would be legitimate consent and the relevant requirements would be unlikely to be made out. See Autoriteit Persoonsgegevens, *AP: scraping bijna altijd illegaal* (*AP*, 1 May 2024) <https://autoriteitpersoonsgegevens.nl/actueel/ap-scraping-bijna-altijd-illegaal > accessed 2 May 2024. For a criticism of the UK Information Commissioner's Office's position in favour of legitimate consent see Guido Noto La Diega, Edina Harbinja, and Katherine Nolan,'BILETA's Response to ICO's Generative AI First Call for Evidence: The Lawful Basis for Web Scraping to Train Generative AI Models' (*SSRN*, 20 February 2024) <https://ssrn.com/abstract=4814018> accessed 2 May 2024.

[44] Google, "Gemini image generation got it wrong. We'll do better", https://blog.google/products/gemini/gemini-image-generation-issue/, accessed 23 February 2024.

move away from grey areas in terms of copyright,[45] but the data pipeline industry is still incipient and could face significant liability as it matures. Concerning deployment, models are often updated with novel training data, which may incur in two phenomena. First, data drift, when the more recent data becomes significantly different from the data used in the original training, increasing model unpredictability. Second, echoes, when content produced by GenAI models is used to train them, creating a feedback loop that can further reinforce existing issues such as bias or hallucinations.

These features and vulnerabilities in GenAI explain why harms have already been occurring, with the pressing need to compensate for the relevant damage, in areas such as (i) discrimination, hate speech, and exclusion; (ii) information hazards; (iii) misinformation harms; (iv) malicious uses; (v) human–computer interaction harms; (vi) environmental and socioeconomic harms.[46] While techniques to ensure ex-ante safety will multiply, there is nothing to suggest that we will be able to design out all harms; therefore, responsible AI will never be meaningful if we do not ensure that the legal systems will react promptly and adequately once these harms materialize. This is the core function of extra-contractual liability.

ENTER THE AI LIABILITY DIRECTIVE: A PRAGMATIC APPROACH TO AI SAFETY?

As recently observed,[47] the US, China, and the EU are dominating the technological space in three different ways, at times converging, at times colliding. These are the presence of most of the top tech companies, increasing control of the infrastructure, and regulation, respectively. Regardless of whether the 'Brussels effect'[48] epitomized by the EU General Data Protection Regulation (GDPR) was overstated,[49] there is no doubt that the EU sees itself as entrusted with the mission to regulate—some may say over-regulate[50]—emerging technologies, and this can be observed with clarity through the lens of AI governance.

The AI Act is an attempt to extend the GDPR model beyond data protection as it purports to apply the new product-safety-like rules to most AI systems, regardless of the geographic location of the supply chain.[51] Unlike the GDPR, however, the AI Act does not venture into liability territory, and this could ultimately limit its influence. At the core of the Act, there is the idea that AI safety can be achieved through ex-ante assessments of the risk associated with an AI system, ie the 'combination of the probability of an occurrence of harm and the severity of that harm'[52]. Ex-ante product safety, however, has never ensured ex-post security, otherwise, no litigation would ever arise to address harms under tortious or contractual liability. This was already the

[45] Katie Paul and Anna Tong, "Inside Big Tech's underground race to buy AI training data", Reuters, https://www.reuters.com/technology/inside-big-techs-underground-race-buy-ai-training-data-2024-04-05, accessed 29 April 2024.

[46] Laura Weidinger and others, 'Taxonomy of Risks Posed by Language Models', *2022 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2022) <https://dl.acm.org/doi/10.1145/3531146.3533088> accessed 18 June 2024.

[47] This is the main thesis of Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (Oxford University Press, Oxford (UK) 2023).

[48] Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press, Oxford (UK) 2020).

[49] For example, the Chinese approach to data protection may be attributed to the Brussels effect as there are many similar provisions in China's Personal information Protection Law (PIPL) and the GDPR; at a closer look, the similarities may have other explanations, and the difference may be deeper than they seem, as convincingly argued by Wenlong Li and Jiahong Chen, 'From Brussels Effect to Gravity Assists: Understanding the Evolution of the GDPR-Inspired Personal Information Protection Law in China' (2024) 54 Comp Law Security Rev 105994.

[50] We doubt the internet ever was a lawless space as the dominant narrative claims, but even if it were there is no doubt that is has now become one of the most heavily regulated sectors, one characterized by complex multi-level and multi-jurisdiction overlaps. See Chris Reed and Andrew Murray, *Rethinking the Jurisprudence of Cyberspace* (Edward Elgar Publishing, Cheltenham (UK) 2018).

[51] For example, the AI Act will apply to AI systems marketed in the EU regardless of where the provider is established, and to the deployers as long as the output produced by the AI system is used in the EU (art 2(1)).

[52] AI Act, art 3(2).

case in a pre-AI world. AI only exacerbates the need to shift the focus from ex-ante actions to ex-post reactions, as it makes it more difficult to predict the occurrence of harms and to fully appreciate their wider repercussions.[53] Shifting the focus also means that liability rules need to be carefully calibrated and harmonized to tackle the issues in AI as presented in the previous section.

Efforts to harmonies European private law have traditionally focussed on contractual law, as epitomized by the Principles of European Contract Law,[54] the UNIDROIT Principles of International Commercial Contracts,[55] and, to a large extent, the Draft Common Frame of Reference.[56] While that path has suffered some setbacks, the often forgotten harmonization of tort law—pushed by the European Group on Tort Law and its Principles of European Tort Law,[57] but hitherto ignored by the EU lawmaker[58]—may lead to a revival of the vision of a European Civil Code. Before the current wave of digital single market laws, the most significant harmonization instrument targeted at non-contractual liability[59] was the Product Liability Directive,[60] aimed at compensating death, personal injury or damage to property caused by a defective product, while alleviating consumers from having to identify a contractual relationship or a fault. Key to its success was that it instantiated a strict liability framework i.e. one where the claimant did not have to prove that the defendant was at fault (e.g. that the defendant was in breach of a duty of care).[61] The current attempt to harmonize non-contractual liability rules is centred on the revision of the Product Liability Directive[62] and on the draft AILD. We will focus on the latter for a threefold reason: the former has received wider attention,[63] it deals with a field of law that has so far been seldom litigated in Europe,[64] and, crucially, it only applies if a harm is

[53] Kretschmer and others (n 12).

[54] Ole Lando and Hugh Beale (eds), *The Principles of European Contract Law. Parts I and II* (Kluwer Law International, The Hague (NL) 1999).

[55] Originally published in 1994 under the stewardship of Michael Joachim Bonell, they have now reached their fourth edition: International Institute for the Unification of Private Law, *UNIDROIT Principles of International Commercial Contracts 2016* (UNIDROIT 2016).

[56] Christian von Bar, Eric Clive, Hans Schulte-Nölke (eds), *Principles, Definitions and Model Rules of European Private Law: Draft Common Frame of Reference (DCFR)* (Sellier, Munich (DE) 2009). The DCFR does deal with non-contractual liability at Book VI, but the vast majority of it deals with contracts.

[57] European Group on Tort Law, *Principles of European Tort Law (PETL)* (EGTL 2005) <http://www.egtl.org/petl.html> accessed 2 May 2024.

[58] This is not to say that the PETL have been irrelevant; eg, they have been repeatedly referred to at the highest jurisdictional levels, eg Italy's Supreme Court, III chamber, no 15536.2017 (in *Resp. civ. Prev.*, 2018, IV, 1148, with comment by Luca Nivarra) about the *compensatio lucri cum damno* referred to the PETL, art 10:103.

[59] Other relevant instruments are Commission Directive 2009/138/EU of 25 November 2009 on the taking-up and pursuit of the business of insurance and reinsurance; Regulation 1215/2012/EU of 12 December 2012 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters; Council Directive 2004/80/EC of 29 April 2004 relating to compensation to crime victims; European Parliament and Council Directive 2004/35/CE of 21 April 2004 on environmental liability with regard to the prevention and remedying of environmental damage. Mauro Bussani and Marta Infantino, 'Harmonization of Tort Law in Europe', in Jürgen Backhaus (ed.) *Encyclopedia of Law and Economics* (Springer, New York City (NY) 2014), 5 refer also to the GDPR's predecessor; while there is no agreement as to the nature of liability for GDPR violations, in England and Wales as well as in the Republic of Ireland there is case law that subsumes it under tort law (*Google v Vidal Hall* [2015] EWCA Civ 311; *Murphy v Callinan* [2018] IESC 59).

[60] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [1985] OJ L 210/29.

[61] Strict liability narrowly construed is based only on causation, typically with a defence for *force majeure* (see Christiane Wendehorst, 'Strict Liability for AI and Other Emerging Technologies' (2020) 11 J Eur Tort Law 150). It is beyond the scope of this contribution to take a stance as to the controversial nature of product liability. See eg, Valeria Pompeo, 'Il Danno Da Prodotti Difettosi' in Paolo Cendon (ed), *Trattato dei nuovi danni. Danni da inadempimento. Responsabilita' del professionalista. Lavoro subordinato*, 4 (CEDAM 2011) 221.

[62] European Parliament legislative resolution of 12 March 2024 on the proposal for a directive of the European Parliament and of the Council on liability for defective products (Second Product Liability Directive) (COM(2022)0495—C9-0322/2022—2022/0302(COD)).

[63] See eg, Guido Smorto and Rosario Petruso, 'Responsabilità Delle Piattaforme Digitali e Trasformazione Della Filiera Distributiva Nella Proposta Di Direttiva Sui Prodotti Difettosi' [2023] Danno e responsabilita' 8, 8; Teresa Rodríguez de las Heras Ballell, 'The Revision of the Product Liability Directive: A Key Piece in the Artificial Intelligence Liability Puzzle' (2023) 24 ERA Forum 247.

[64] We predicted that the rise of the Internet of Things might lead to a revival of product liability litigation in Guido Noto La Diega and Ian Walden, 'Contracting for the "Internet of Things": Looking into the Nest' (2016) 7 Eur J Law Technol <http://ejlt.org/article/view/450> accessed 30 April 2019. We would accept that the prediction did not prove to be accurate.

the consequence of a defect in a product. Conversely, the proposed AILD constitutes the first attempt by the EU lawmaker to make tangible progress towards the horizontal harmonization of general tort law. As such, it deserves closer inspection.

After the launch of the *European strategy for AI* in 2018,[65] the Expert Group on Liability and New Technologies published a report on liability for AI and other emerging digital technologies in 2019.[66] There, it observed that, while domestic liability regimes ensure basic protection of victims of AI-caused damage, the characteristics of these technologies (e.g. modification through self-learning during operation, limited predictability, etc.) and their applications may make it more difficult to compensate victims, and in some scenarios the allocation of liability would be unfair or inefficient. Accordingly, they concluded that 'certain adjustments need to be made to EU and national liability regimes.' Building on this, in 2020, the European Commission published its *White Paper on AI*[67] and accompanying report,[68] tasking the EU with the goal of setting forth a framework to tackle some risks associated to AI as a way to promote the technology's development and adoption. The *White Paper* proposed to focus on three interrelated issues:

(i) The development of a horizontal regulatory framework for AI, focussing on issues of safety and fundamental rights—the AI Act would become its centrepiece;
(ii) The revision of existing sectoral safety legislation, with the new General Product Safety Regulation[69] and the new Machinery Regulation[70] as its cornerstones; and
(iii) Updated rules on AI liability, which would ultimately lead to the proposed Second Product Liability Directive and the draft AILD.

To justify the need for a reform of liability rules, the Commission underlined the heightened likelihood of harm and the difficulty to apportion liability due to the integration of AI into products, design flaws, poor quality or availability of data, and limited access to evidence. Consequently, despite the presence of ex-ante product safety laws, 'if the safety risks materialize, the lack of clear requirements and the characteristics of AI technologies […] make it difficult to trace back potentially problematic decisions made with the involvement of AI systems [as well as making it difficult] for persons having suffered harm to obtain compensation under the current EU and national liability legislation.'[71] Adding to the momentum, the European Parliament called on the Commission to propose legislation on civil liability for AI,[72] resulting in the *2021 Coordinated Plan for AI* that articulated the objective to introduce EU measures adapting the liability framework to the challenges of new technologies, including AI, and expressly stating that the new framework would include 'a revision of the Product Liability Directive, and *a legislative proposal with regard to the liability for certain AI systems*.'[73]

[65] European Commission, *Communication "AI for Europe"*, COM/2018/237 final.
[66] Expert Group on Liability and New Technologies—New Technologies Formation, *Liability for Artificial Intelligence and Other Emerging Digital Technologies* (EU 2019) 3.
[67] European Commission, 'White Paper on Artificial Intelligence – A European Approach to Excellence and Trust COM(2020) 65 final.
[68] European Commission, 'Report on the Safety and Liability Implications of Artificial Intelligence, The Internet of Things and Robotics' (2020) COM 64 final.
[69] Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council and Directive (EU) 2020/1828 of the European Parliament and the Council, and repealing Directive 2001/95/EC of the European Parliament and of the Council and Council Directive 87/357/EEC [2023] OJ L 135/1.
[70] Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC [2023] OJ L 165/1.
[71] European Commission, *White Paper on Artificial Intelligence* (n 55) 12.
[72] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)) [2021] OJ C 404/107.
[73] European Commission, *Coordinated plan on artificial intelligence 2021 review,* Annex to the Communication 'Fostering a European approach to Artificial Intelligence' (2021)COM 205 final, 33.

In September 2022, the Commission followed up with the proposed AILD, which as we write is going through its first reading.[74] As the EU lacks general competence to fully harmonize tort law,[75] Article 114 TFEU provides the legal basis for the proposal; indeed, this measure is seen as pivotal to ensuring the good functioning of the internal market by removing the legal uncertainty and fragmentation that hinders cross-border trade in AI-powered goods and services.[76] While harmonization has its costs and therefore its need must be justified[77], an economic study has shown that, when it comes to AI liability, uniform liability rules are set to have a positive impact of 5–7% on the production value of relevant cross-border trade.[78] The Commission has adopted a 'staged approach'[79] where the AILD would introduce some minimum harmonization measures[80] to ease claimants' burden of proof in scenarios where the harm is due to fault (typically a breach of a duty of care), whereas in five years there would be a re-assessment of the need for strict liability, possibly coupled with mandatory insurance. The following subsections present a concise critical appreciation of the three pillars of the draft AILD: (i) subject matter, scope, and definitions; (ii) disclosure of evidence in cases involving high-risk AI systems; (iii) presumption of causal link between fault and damage caused by any AI system.

Subject matter, scope, and definitions in the AILD

A major drawback of the draft AILD is that it has a significantly narrow scope, namely certain aspects of non-contractual fault-based civil liability for damage caused by AI systems.[81] 'Certain aspects' means that the harmonization is limited to (i) the disclosure of evidence to enable a claimant[82] to substantiate a non-contractual fault-based civil law claim for damages caused by a high-risk AI system;[83] and (ii) the burden of proof in the event of non-contractual fault-based civil law claims brought before national courts for damages caused by an AI system. This instrument is set to work in parallel to the new Product Liability Directive, with the key differences that the former applies only to intangible AI systems,[84] the latter to both defective physical products and defective software[85]. As a growing number of objects becomes embedded with AI and as only some harms will be the consequence of a defect, this choice is open to criticism. To say that the proposed Directive introduces '*very targeted* and proportionate'[86] measures to ease the burden of proof through disclosure and rebuttable presumptions may be an understatement, and it begs the question of whether this excessive 'targeting' may be framed as a lack of ambition.

The scope—and accordingly the potential to impactfully rewrite the rules of AI liability in Europe—has several additional limitations. First, the proposed AILD is a minimum harmonization measure i.e. national laws could remain in place if more favourable for claimants.[87] This

[74] As we write, the proposal has just been referred to the Internal Market and Consumer Protection (IMCO) committee and the Civil Liberties, Justice and Home Affairs (LIBE) committee.

[75] Bussani and Infantino (n 62) 5.

[76] Draft AILD, recitals 6, 7, 9, 12, 32; Draft AILD Explanatory Memorandum, para 2.

[77] Michael G Faure, 'Product Liability and Product Safety in Europe: Harmonization or Differentitation' (2000) 53 Kyklos 467.

[78] Deloitte, *Study to Support the Commission's IA on Liability for Artificial Intelligence*, 2021.

[79] Draft AILD Explanatory Memorandum [1].

[80] Draft AILD, recital 14.

[81] Draft AILD, art 1(2); Draft AILD Explanatory Memorandum, 11.

[82] Alongside the individual injured person, a successor (e.g. the victim's heir, the insurance company, etc.), and the actor in a class action (Draft AILD, arts 2(6) and 6, the latter inserting point 67 in Directive (EU) 2020/1828 of the European Parliament and of the Council of 25 November 2020 on representative actions for the protection of the collective interests of consumers and repealing Directive 2009/22/EC [2020] OJ L 409/1, Annex I).

[83] Draft AILD, art 1(1)(a); see s 3.2 below.

[84] Draft AILD, recitals 7 and 11; European Commission, Questions and answers on the revision of the Product Liability Directive (2022) [9].

[85] This is a departure from the text that the European Parliament recommended in its resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)), see first Annex, art 3(a)(b).

[86] Draft AILD Explanatory Memorandum, 11.

[87] Draft AILD, art 1(4).

could end up being beneficial for some claimants. For example, Member States could maintain national strict liability regimes[88] eg Italy could keep its liability regime for dangerous activities under Article 2050 of the *Codice civile*. This provision arguably applies to AI damages, especially now that the AI Act provides guidance as to what systems are high-risk[89]. Under Article 2050 *codice civile*, '[c]ompensation must be paid by whomever damages others while exercising an activity that is either dangerous by its very nature or due to the means used, unless the defendant can prove to have put in place suitable measures to prevent the damage'[90]. This regime is particularly useful in the context of damage caused by 'technological unknown', ie damages that could not be predicted in light of the state of the art in the relevant technological field[91]. It is praiseworthy that more protective regimes will remain in place, even though this applicability is subject to the more favourable being 'compatible with Union law'[92], which is an opaque condition. Importantly, it is a missed opportunity not to harmonize those regimes that are more directly relevant when it comes to AI damages, such as the aforementioned liability for dangerous activities, which could constitute the model upon which the EU law of AI torts is built. It has been argued that minimum harmonization is a good choice as it allows Member States to experiment with different rules, which may be 'beneficial at the beginning of the deployment of a new category of technologies'[93]. This position is exposed to a twofold criticism. First, it ignores the current function of the regulation of technologies (and the harmonization of torts has a clear regulatory function), ie a tool to assert one's power internationally[94]. Second, a 70-year-old technology that is now living its most lively phase can hardly be treated as something new that should not be governed organically otherwise we risk killing innovation.

The proposed AILD leaves also criminal liability[95] out of the scope and would not affect Union laws on liability in the field of transport,[96] product liability,[97] intermediary liability as recently reformed by the Digital Services Act,[98] as well as national rules on who carries the burden of proof, the relevant standard, and the definition of fault. This final exclusion comes with the condition 'other than in respect of what is provided in Articles 3 and 4'[99], which we turn our attention to in the next sub-sections. For now, suffice it to say that by curtailing its scope in such a drastic way, albeit at times justified, the EU lawmaker may have reduced the AILD's significance and prevented its ultimate success. In the *Explanatory Memorandum,* one can read that the proposal avoids touching on the definition of fundamental concepts e.g. fault or damage '*given that the meaning of those concepts varies* considerably across Member States'[100]. While we

[88] Draft AILD, recital 14.

[89] On the applicability of this regime to AI see eg Antonino Procida Mirabelli di Lauro, *Intelligenze artificiali e responsabilita' civile*, in Antonino Procida Mirabelli di Lauro and Maria Feola, *Diritto delle obbligazioni* (ESI 2020) 507, esp. 534 ff. The main limitation is that this regime is alternative to the product liability one. There is also the issue that, despite the AI Act, there remains uncertainty as to what constitutes a high-risk system, as we will note in the following section.

[90] Codice Civile, art 2050.

[91] Lalage Mormile, 'Il principio di precauzione fra gestione del rischio e tutela degli interessi privati', *Riv dir ec trasp amb* 10 (2012): 247, esp 271. Some of the issues related of the development risk defence (also known as the state-of-the-art defence) are being addressed in the revision of the Product Liability Directive, which may to some extent reduce the usefulness of falling back on the liability for dangerous activities.

[92] Draft AILD, art 1(4).

[93] Jan De Bruyne, Orian Dheu and Charlotte Ducuing, 'The European Commission's Approach to Extra-Contractual Liability and AI—An Evaluation of the AI Liability Directive and the Revised Product Liability Directive' (2023) 51 Comp Law Security Rev 105894, 3.

[94] See for example the UK National Cyber Strategy 2022, which—rather than focusing on cybersecurity per se as the previous strategy did—centres on the idea of cyber power as 'the ability to protect and promote national interests in and through cyberspace' (ibid 11), against competitors such as China and Russia.

[95] Draft AILD, art 1(2).

[96] Draft AILD, art 1(3)(a).

[97] Draft AILD, art 1(3)(b).

[98] Draft AILD, art 1(3)(c); Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (DSA) [2022] OJ L 277/1, arts 4-10.

[99] Draft AILD, art 1(3)(d).

[100] Draft AILD Explanatory Memorandum, 11, emphasis added.

do not wish to downplay the resistance that a more holistic intervention may cause, harmonization intervenes precisely to tackle the variance in concepts that is at odds with the needs of the internal market. This variance calls for harmonization measures, rather than justifying their lack. Additionally, one could call into question whether it is true or even possible to legislate AI liability without dealing with key questions of fault and damage. Indeed, the draft AILD defines a 'claim for damages' as one that compensates 'damage caused by an output of an AI system or the failure of such a system to produce an output where such an output should have been produced.'[101] Leaving aside the lack of clarity as to when it can be said that an output should have been produced, the quoted provision is nothing less than an incomplete and indirect way of defining damage. In the section related to Article 4 AILD we will see that the same can be said of the concept of fault.

A bolder and more comprehensive approach would have been desirable. The Commission explains that they did not go ahead with more far-reaching changes such as a full reversal of the burden of proof and strict liability because '*businesses provided negative feedback* in consultations'[102] and that the AILD therefore appeared as a more pragmatic, targeted, and proportionate response. However, one can question what is the function of regulating technologies, if not to incentivize private business to behave responsibly, as otherwise their sole incentive would be to maximize profit and the harm to citizens and society would be for the most part ignored as a mere externality.[103] The opposition from AI companies to more protective reforms demands ambitious legislative intervention, as opposed to justifying its absence. More generally, we posit that, to give responsible AI due consideration, we need a much more comprehensive approach to the harmonization of torts. If we do not fully harmonize tort law—starting with AI, which in the long run is likely to become the hidden infrastructure of the world—we risk defeating the purpose of responsible AI and AI safety. It has been put forward that '[i]t is unrealistic to harmonize specific liability rules for damage caused by AI systems without reaching an agreement on a common liability framework beforehand'[104]. We disagree for two reasons. First, in the long run, as AI becomes weaved into the fabric of our society, virtually all harms will be AI harms. Second, for years general tort harmonization projects have been presented and failed, instead one could adopt the incremental approach that is typical in technology development: new technological tools—and new legal frameworks—have to be introduced gradually, thus leaving open the possibility to step back when unintended consequences occur[105]. AI could provide the opportunity to test the harmonization of torts and extend it gradually to become technologically neutral. Much is at stake, if general private law does not take its regulatory role seriously. All focus risk shifting to the various technology-specific or market-specific regulatory interventions (AI Act, Digital Services Act, etc.)[106] and general private law could be perceived as the irrelevant residue of a world long gone.

Disclosure of evidence in cases related to high-risk AI systems

While quantity is not necessarily determinative of quality, the fact that the proposed AILD is extremely succinct and it contains only two substantial provisions can be regarded as an

[101] Draft AILD, art 2(5).
[102] Ibid, emphasis added.
[103] Along similar lines, Kayleen Manwaring, 'Will Emerging Technologies Outpace Consumer Protection Law? The Case of Digital Consumer Manipulation' [2018] Competition Consumer Law J 141.
[104] Li and Schütte (n 22) 151.
[105] In this sense, though in the context of educational AI, see European Commission Expert Group on AI in Education, *Final report of the Commission expert group on artificial intelligence and data in education and training* (European Union 2022) <https://op.europa.eu/en/publication-detail/-/publication/7f64223f-540d-11ed-92ed-01aa75ed71a1/language-en> accessed 21 June 2024.
[106] See Vagelis Papakonstantinou and Paul De Hert, *The Regulation of Digital Technologies in the EU: Act-Ification, GDPR Mimesis and EU Law Brutality at Play* (Routledge 2024).

indicator that the EU is far from taking the task of harmonizing torts seriously. Let us not allow such paucity to prevent us from analysing both in turn.

Article 3 focuses on the key issue of discovery: it empowers national courts to request the disclosure of evidence, and in some instances its preservation. Claimants can be granted disclosure orders provided that they (i) address one of the persons expressly listed in the provision (the provider of high-risk AI system, the product manufacturer, the user, etc.)[107]; (ii) identify a specific high-risk AI system that is suspected of having caused a damage; (iii) made all proportionate attempts at gathering evidence from the defendant; (iv) present elements to corroborate the plausibility of a claim for damages.[108] Potential (as opposed to current) claimants have to meet the additional requirement of demonstrating that they had asked the prospective recipient to disclose evidence, and the latter had refused.[109] Additionally, potential claimants are not entitled to preservation orders.[110] As elaborated in the recitals, the rationale of the provision is that access to information is instrumental to deciding about the viability of a potential lawsuit, to substantiate claims for compensation, and to complement the AI Act's documentation, information and logging requirements[111].

At closer inspection, three reasons make us say that the change is less meaningful than one would have hoped. The first one relates to the aforementioned requirements, the other two are more general considerations. Starting off, the requirements for a discovery or preservation order to be granted clearly ignore the 'many hand' problem of the complex supply chain of AI[112], the notorious black box issue[113]—both exacerbated in the GenAI context—and the increased power imbalance[114] that stems from this opacity as well as from the legal and techno-factual control AI players retain over information. Opacity and the many hands problem mean that potential and actual claimants may struggle to find out who a suitable recipient for the discovery or preservation order might me[115]. It would also mean that they would find it difficult to identify a specific high-risk AI system that would likely have caused damage. The control that AI players retain over data could be used to frustrate any attempts at gathering evidence from the defendant, as well as to prevent the claimant or prospective claimant from being able to gather elements to corroborate the plausibility of a claim for damages.

The second cause for caution is that the provision only applies to systems that the AI Act regards as posing a high risk to health, safety, and fundamental rights. Computer science scholarship has pointed out how classifying an AI system as 'high-risk' is a complex endeavour.[116] One that is not helped by the EU lawmaker, as the AI Act fails to adequately define what it means by 'high-risk

[107] The reference is to the providers of high-risk systems, the user, and 'a person subject to the obligations of a provider pursuant to [Article 24 or Article 28(1)' of the AI Act in the originally proposed version, i.e. certain product manufacturers (e.g. toys, lifts, etc.) and, under certain circumstances, distributors, importers and third parties (e.g. if they make a substantial modification to the high-risk system). See arts 23-26 of the final version of the AI Act; as the adopted version differs significantly from the proposal that the draft AILD referred to, it is expected that significant work will be needed to avoid misalignments.

[108] Draft AILD, art 3(1) and (2).

[109] Draft AILD, art 3(1).

[110] Draft AILD, art 3(3).

[111] Draft AILD, recital 6.

[112] See eg Donal Khosrowi, Finola Finn and Elinor Clark, 'Engaging the Many-Hands Problem of Generative-AI Outputs: A Framework for Attributing Credit' [2024] AI and Ethics <https://doi.org/10.1007/s43681-024-00440-7> accessed 16 May 2024.

[113] More on this already in Guido Noto La Diega, 'Against the Dehumanisation of Decision-Making—Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information' (2018) 9 JIPITEC 3.

[114] Jorge Luis Morton Gutiérrez, 'On Actor-Network Theory and Algorithms: ChatGPT and the New Power Relationships in the Age of AI' [2023] AI and Ethics <https://doi.org/10.1007/s43681-023-00314-4> accessed 16 May 2024.

[115] The draft AILD itself recognizes that '[t]he large number of people usually involved in the design, development, deployment and operation of high-risk AI systems, makes it difficult for injured persons to identify the person potentially liable for damage' (recital 17). The same applies to the identification of the discovery order recipients.

[116] Delaram Golpayegani, Harshvardhan J Pandit and Dave Lewis, 'To Be High-Risk, or Not To Be—Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards', *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2023) <https://dl.acm.org/doi/10.1145/3593013.3594050> accessed 13 May 2024.

AI system'. Rather, it provides a list of systems that are regarded as high-risk (eg e-proctoring software)[117] and introduces a set of exemptions (eg AI used to detect patterns in decision-making are not high risk) and then an exception to the exemptions, ie the exempt systems will be regarded as high risk if used for profiling purposes.[118] To add to the complexity, AI providers can disagree with the classification if they believe that the system is in fact not high risk. Positively, predictability may be increased by the guidelines that the Commission is set to adopt; they should provide examples of high-risk and non-high-risk use cases.[119] Nonetheless, legal certainty is under threat as the criteria for classifying systems will change over time under yet-to-be adopted delegated acts.[120]

The third drawback of Article 3 has to do with its provisions subjecting the grant of the disclosure and preservation orders to a necessity and proportionality assessment. To understand the issue, it can be helpful to re-frame the provision in the context of the ongoing US-EU legal catch-up. Indeed, Article 3 can be construed as tackling the general problem of the lack of harmonisation of civil law discovery rules in Europe, and their weakness vis-à-vis US discovery rules. It is widely accepted that '[w]ide ranging pretrial discovery is an integral part of contemporary American civil litigation',[121] with European—and more generally civil law—discovery being much more limited. The most commented-on difference is that access in Europe will be curtailed under data protection law.[122] Under US law, the main limitation is that privileged information may be withheld.[123] At first glance, it would seem that the EU is modelling the reform on the US as it requires national courts to evaluate whether the disclosure or preservation would harm any trade secrets and confidential information.[124] In fact, the European approach remains weaker—and surprisingly more proprietary than the US—for at least two reasons. First, European courts will be required to consider not only trade secrets and confidentiality, but also 'the legitimate interests of all parties, including third parties'[125]. This seems much broader than the concept of privilege.[126] Second, the proposed AILD introduces measures safeguarding the AI company as opposed to the victim, namely (i) evidence will only be disclosed if necessary and proportionate to support a claim or potential claim; (ii) when upon assessment discovery is granted, national courts will order measures to preserve confidentiality; (iii) EU Member States will be required to change their laws to introduce appropriate remedies against disclosure or preservation orders.[127] By contrast, in the US the safeguards are focussed on making sure that the defendant does not abuse the control they hold over the information. Notably, the party wanting to withhold information has to make an express claim that the information is privileged and, crucially, they have to detail the nature of undisclosed information so as to enable the other parties to assess the claim.[128] Lacking these safeguards, there is the risk that in Europe defendants will abuse their IP or de facto control over the information to neutralize the new provisions. This is in line with research that claims that the new digital single market laws' emphasis on

[117] Draft AI Act, art 6(1)-(2) and Annex III [3](d).
[118] Draft AI Act, art 6(3).
[119] Draft AI Act, art 6(5).
[120] Draft AI Act, art 6(7).
[121] Geoffrey C Jr Hazard, 'Discovery and the Role of the Judge in Civil Law Jurisdictions' (1997) 73 Notre Dame Law Rev 1017.
[122] Carla L Reyes, 'The U.S. Discovery-EU Privacy Directive Conflict: Constructing a Three-Tiered Compliance Strategy Note' (2008) 19 Duke J Comp Int Law 357; Eckard von Bodenhausen, 'U.S. Discovery and Data Protection Laws in Europe' (2012) 37 DAJV Newsletter 14.
[123] 29 CFR § 18.51(e). With regards to the assistance to foreign and international tribunals, see 28 USC §1782.
[124] Draft AILD, art 3(4).
[125] Draft AILD, art 3(4).
[126] Privilege refers mostly to client-attorney privilege (*Upjohn Co. v. United States* 449 U.S. 383 (1981)), attorney work-product (*United States v. Nobles* 422 U.S. 225, 238–39 (1975)), and joint defence privilege (*United States v. Henke* 222 F.3d 633 (9th Cir. 2000)), with some US states recognizing other privileges, eg between ministers and their confessors (eg Cal. Evid. Code § 912).
[127] Draft AILD, art 3(4).
[128] 29 CFR § 18.51(e).

striking a balance between data access and IP can be easily exploited and lead to overprotection of IP and ultimately closed, inscrutable AI[129].

On a more positive note, the harmonized discovery provision is equipped with an incentive for the defendant to comply with the discovery or preservation order. Indeed, if they fail to do so, national courts shall presume the non-compliance with the duty of care that the withheld evidence was intended to prove. This presumption is limited to the defendant not complying with the disclosure/evidence order; it will be of no help if the evidence is held by other AI players, which may limit its usefulness. An amendment to the proposed AILD to ensure its extra-territorial application—similar to the AI Act, the GDPR, etc.—would resolve this problem, at least in part, as it would ensure that the provider, manufacturer, etc. could become defendants regardless of where they are established. The defendant can rebut this presumption, which might have the perverse outcome of defendants waiting until after the judicial application of the presumption to disclose the relevant evidence. National courts will have to be careful in administering this mechanism in way that does not allow for dilatory tactics.

Presumption of causal link between the fault and the damage caused by the AI

The most important innovation on the presumptions front is introduced by Article 4, that accounts for the difficulties to prove the causal link between the damaging AI output (or lack thereof) and the defendant's fault. These difficulties stem from the intrinsic features of most AI systems, notably autonomy and opacity[130], as seen in a previous section. To account for them, the provision empowers national courts to presume the existence of the causal link, while leaving it to the defendant to rebut the presumption. For the presumption to operate, three cumulative conditions must be met:

(i) The claimant has demonstrated—or the judge has presumed under Article 3—the defendant's fault i.e. the non-compliance with a duty of care 'directly intended to protect against the damage'[131];
(ii) The circumstances make it 'reasonably likely'[132] that the fault did influence the output or its lack;
(iii) The claimant has proved that the damage derived from the output or lack thereof.

The proposed AILD goes to great lengths to make it clear that it does not intend to harmonize the concept of fault or the conditions under which domestic courts establish fault.[133] It must be questioned whether it is at all possible to harmonize the rules about the causal link between fault and output without intervening on the concept of fault. In fact, we would argue that the draft AILD can be interpreted as a backdoor reform of civil fault, which is set to acquire an autonomous meaning in EU law, ie 'a human act or omission which does not meet a duty of care under Union law or national law that is directly intended to protect against the damage that occurred'[134]. This is much narrower, compared to most national conceptions of fault. Taking the Italian legal system as an example, the *colpa* is seen as the expression of the general *alterum non laedere* principle (do not harm), a standard[135] that arises in the event of negligence, recklessness,

[129] Guido Noto La Diega, 'Ending Smart Data Enclosures: The European Approach to the Regulation of the Internet of Things between Access and Intellectual Property' in Stacy-Ann Elvy and Nancy Kim (eds), *The Cambridge Handbook on Emerging Issues at the Intersection of Commercial Law and Technology* (Cambridge University Press 2024) 258.
[130] Draft AILD, recital 27.
[131] Draft AILD, art 4(1)(a).
[132] Draft AILD, art 4(1)(b).
[133] For example draft AILD, recitals 22–23; Explanatory Memorandum, 11.
[134] Draft AILD, recital 22; and nearly verbatim art 4(1)(a).
[135] Mario Barcellona, *La responsabilità civile*, in Salvatore Mazzamuto (dir), *Trattato del diritto privato,* vol 6 (Giappichelli 2021) 165.

incompetence, or illegality[136]. The European concept would be limited to the fourth type of *colpa;* fault by 'illegality' stems from the non-compliance with those legal provisions setting forth measures aimed at avoiding or minimizing the risk of harm[137]. This notion of fault is additionally curtailed; indeed, not all safety rules would be relevant, but only the duties of care that were 'directly intended to protect against the damage that occurred'.[138] For example, the non-compliance with the AI Act's documentation requirements—requirements whose compliance the draft AILD is designed to incentivize—would not lead to the application of the causal link presumption.[139] Conversely, there would be fault in the event of physical injury that is the consequence of the ignorance of the instructions of use which are specifically designed to prevent harm to natural persons. People harmed by AI and courts will have to carry out a delicate exercise of identifying those Union and national laws that would fall within this narrow scope. This task is easier when it comes to the high-risk AI systems, as the proposed AILD sets forth a brief and exhaustive list of provisions whose breach can lead to fault (eg data quality rules)[140]. While the provision of this list is positive from the standpoint of legal certainty, it can be criticized from a responsible AI perspective as it further hampers the impact and relevance of the instrument.

Due to the narrow concept of fault, there is the risk that reductive interpretations of Article 4 will prevail, thus undermining the do not harm principle that is at the basis of the laws of robotics.[141] This is further confirmed by three additional limitations to the presumption of a causal link; these can be regarded as hidden additional requirements, on top of the three requirements mentioned at the beginning of this section. First, if the defendant proves that 'sufficient evidence [about the causal link] and expertise is reasonably accessible'[142] the court will not apply the presumption. Limited to harms caused by high-risk systems, this limitation is aimed at incentivizing the defendant to comply with their disclosure obligations, including transparency measures mandated by the AI Act[143]. While understandable, this limitation would have the perverse effect of treating providers of high-risk systems more favourably than providers of medium and low-risk systems. Second, another hidden requirement is buried in one of the final paragraphs of Article 4. Namely, the national court must be satisfied that it would be 'excessively difficult'[144] for the claimant to prove the link. From a legislative drafting point of view, it would have been preferable to include this limitation as one of the requirements in the opening of the provision. In terms of policy, this seemingly high bar may mean that the presumption can be invoked only in a minimal number of cases. Third, the presumption will apply only if a defendant who used the AI system for personal, non-professional purposes 'materially interfered with the conditions of the operation of the AI system or if [they were] required and able to determine the conditions of operation of the AI system and failed to do so'[145]. Much will depend on what will constitute a material interference, with the added uncertainty related to use by prosumers. Considering

[136] Italy's Civil Code, arts 2043 and 1176(1) c.c. and Criminal Code, art 43(1); see the monographic analysis conducted by Laura Mancini, *La colpa nella responsabilità civile* (Giuffrè 2015).

[137] C Massimo Bianca, *Diritto civile. La responsabilità*, vol 5 (2nd edn, Giuffrè 2019) 579.

[138] Draft AILD, recital 22.

[139] Draft AILD, recital 22.

[140] With regard to claims for damages against a provider and those who are subject to the same obligations under certain circumstances (eg, manufacturers), the draft AILD, art 4(2) refers also to certain transparency, human oversight, accuracy requirements and the provisions on corrective actions to bring the system in line with certain obligations under the AI Act. When it comes to claims against the user, the claimant needs to prove the non-compliance with the AI Act's obligations to use or monitor the system in line with the instructions, or not to expose the system to input data that is not relevant in view of the system's intended purpose (art 4(3)).

[141] The reference is to Asimov's three laws of robotics, see Ugo Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts* (Springer, Dordrecht (NL) 2013).

[142] Draft AILD, art 4(4).

[143] Draft AILD Explanatory Memorandum [4].

[144] Draft AILD, art 4(5).

[145] Draft AILD, art 4(6).

these drawbacks, one cannot but wonder why the legislator—rather than introducing a total of six requirements for the presumption to apply—did not simply provide that the defendant retains a right to rebut the causal link presumption. Arguably a much neater solution, one that would be more in line with the policy goals of this instrument.

Finally, the uncertainties related to the concept of fault and the burden of proof that remains heavy despite the discovery rules and the presumption of causal link create a mismatch between the complexity of the law and the simplicity required by AI as an automation engine. In the next section, while evaluating the AILD's fitness for GenAI, we will argue that this is an argument in favour of a shift to strict liability.

IS THE AI LIABILITY DIRECTIVE FIT FOR AI AT THE TIME OF CHATGPT?

Alongside the intrinsic drawbacks highlighted in the previous sections, there remains the cross-cutting general question of whether the proposed AILD is fit for AI in its current state of development, ie a GenAI-dominated world. To answer this question, one needs to consider (i) the relationship to the AI Act; (ii) the functions of the concept of risk; (iii) whether the features of GenAI and the harms typically ensuing are adequately tackled in the draft AILD. We will argue that strict liability would be the most adequate response to current and predictable AI harms.

The proposed AILD is modelled on the initial version of the AI Act. As we write, the Council has not resumed its work after the European elections. When it does, there will be a great deal of work needed to re-align the two instruments, so much as that—as we will call for—it would be easier to scrap the proposal and start from scratch. The main difference between the draft AI Act and the adopted version is that the former revolved around the concept of risk—with high-risk systems being the target of most safety requirements—and having applied AI of the predictive and decisional types as its focus, as clearly shown by Annex III and the list of applications that would be regarded as high-risk. Instead of pausing and rewriting the AI Act to account for general and generative AI, the EU lawmakers have ploughed ahead by adding several provisions about general purpose AI, including some additional requirements for models with systemic risk[146]. Unlike what the official press releases would suggest[147], it does seem that the risk-based approach has been abandoned and that GenAI provisions are the core of the Act, regardless of the (predictability of the) risk they pose to safety, fundamental rights, and society.[148] This means that it will have to be decided whether to extend the AILD's references to all AI systems within the scope of the AI Act, extend them to all general-purpose AI models, or only to the ones with systemic risk. It does not seem that it would be systematically tenable to leave things as they are i.e. limited to high-risk narrow AI.

Relatedly, but more generally, a deeper reflection should be developed around the point of a risk-based approach as applied to torts, a body of law that, by its very nature, operates after a harm has occurred. Risk-based norms make have a straightforward function when it comes to product-safety-like legislation that operates ex ante. Indeed, the whole point of these laws is to prevent harm from occurring or minimizing the chance of it occurring. Conversely, the

[146] General purpose AI models have system risk when, due to their high-impact capabilities, are or will foreseeably be detrimental to 'public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain' (AI Act, art 3(65)).

[147] 'Artificial Intelligence (AI) Act: Council Gives Final Green Light to the First Worldwide Rules on AI' (*EU Council*, 21 May 2024) <https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/> accessed 17 June 2024.

[148] Indeed, even though there are some additional requirements for general purpose AI with systemic risk, they have been significantly watered down and now revolve around the idea of compliance through codes of practice (arts 55(2) and 56).

concept of risk has a different relevance when it comes to apportioning liability for harms that have already occurred, which is what the AILD would help doing. This is not to say that risk performs no function in the context of torts. At the EU level, it justifies strict liability systems such as liability for defective products[149]. As the Product Liability Directive states, '*liability without fault* on the part of the producer is the *sole means of adequately solving the problem*, peculiar to our age of increasing technicality, of a fair apportionment of the *risks inherent in modern technological production*.'[150] At the domestic level, going back to the Italian extra-contractual liability regime for dangerous activities under Article 2050 of the *Codice civile*, risk justifies once again a form of no-fault liability. This begs the question of whether strict liability should be adopted as the default for all harms caused by an AI. This option would have the following benefits:

(i) Increasing the trustworthiness of AI systems, and with it their uptake;
(ii) Providing strong incentives for the companies responsible for AI development and deployment to rigorously operationalize the safety rules of the AI Act and other safety legislation;
(iii) Accounting for the features of AI and the likely harms as discussed in a previous section.

One could put forward the objection that strict liability does not work for GenAI, namely for foundation models, because '[o]ne Foundation Model, however, might be used in 1000 AI applications, only one of them being a high-risk application.'[151] This point goes hand in hand with the suggestion that—rather than horizontally regulating liability in torts for all AI systems, one should prefer a technology-specific approach 'identifying single classes of applications that need to be separately regulated with independent normative acts'[152]. The objection is not without merit, but for the reasons above we disagree with the idea of a gradation of ex-post liability based on risk. The limited control that the providers of upstream models have on downstream applications can be accounted for in different ways, eg by leaving it to the defendant to prove *force majeure* or fortuitous event[153]. It is true that imposing strict liability on the providers of foundation models can be perceived as a drastic policy option, but we will be soon accepting the notion that the providers of foundation models are providing an essential service, even an essential facility, and that with great power comes greater responsibility—and in some instances also greater liability. Another objection to a strict liability framework for AI is that it would constitute an excessive burden on business and stifle innovation; indeed, the reasoning goes, that strict liability would lead to a dramatic increase in litigation, and this litigation would always be decided in favour of the claimant. Both worries are largely unwarranted. First, EU strict liability rules have been around for a long time, and they have produced a limited number of disputes[154]. Second, strict liability does not mean that there would be no defences available to AI

[149] More on the concept of risk in the Product Liability Directive in Daily Wuyts, 'The Product Liability Directive—More than Two Decades of Defective Products in Europe' (2014) 5 J European Tort Law 1.

[150] Product Liability Directive, recital 2. This is one of the few parts of the directive that remains mostly unchanged in the Second Product Liability Directive (recital 2).

[151] Philipp Hacker, 'The European AI Liability Directives—Critique of a Half-Hearted Approach and Lessons for the Future' (2023) 51 Comp Law Security Rev 105871, 32.

[152] Andrea Bertolini, 'Artificial Intelligence and Civil Liability' (JURI Committee 2020) 87.

[153] The Italian regime of strict liability for dangerous activities leaves it to the defendant to prove that they adopted all safeguards suitable to avoid the harm (Codice Civile, art 2050). A similar regime is provided in under Article 2051 of the Civil Code, which is also likely to be relevant in the context of AI harms, as observed by Barcellona (n 137) 266. Article 2051 applies to the harm caused by the things that where within the defendant's control e.g. the company responsible for managing a dam can be held liable in the event of flooding (Corte di Cassazione, Sezioni Unite, ordinanza No 20943 of 30 June 2022, in CED Cassazione [2022]). The defendant can escape liability by proving that the harm was caused by a fortuitous event.

[154] This may change with the proposed Second Product Liability Directive, which has been rewritten to account for the rise of AI and IoT technologies (recitals 3, 17, 18, 32, 50). However, the instrument contains a number of provisions aimed at 'address[ing] a potential risk of litigation in an excessive number of cases' (recital 22). The proposed directive is currently awaiting the EU Council's first reading position.

companies. Risk, once again, may come in handy as there is no liability for defective products in the event of scientifically unknowable risk.[155] Something along the lines of this defence could be envisaged for AI torts. If one considers the *travaux préparatoires* of the proposed AILD, it becomes immediately apparent why strict liability was disregarded despite having overwhelming support from citizens, consumer groups, and academics[156]. Indeed, the 'majority of business respondents'[157] considered the no-fault policy option to be disproportionate. As this instrument is meant to contribute a single market for AI in Europe, it is no surprise that the EU lawmaker would give weight to the views of private business. However, these views should not be the deciding factor, especially considering that the economic study that the proposed AILD is based on states that 'the moderate compliance costs linked to [the strict liability policy option] would be outweighed by cost savings thanks to higher legal certainty, saved resources on compliance, and higher revenue enabled by a clearer and less fragmented legal framework.'[158] Additionally, a strict liability framework could be complemented by measures to reduce the compliance burden for companies, eg liability caps and subsidized insurance for SMEs[159]. As the problems that the AILD was designed to tackle were legal uncertainty, legal fragmentation, and lack of compensation leading to mistrust and limited uptake of AI, the policy option of minimum harmonization fails across the board[160], and risks penalizing SMEs in particular[161]. Article 5 does require the Commission to evaluate the impact of the AILD in five years, including legislative proposals to introduce 'no-fault liability rules for claims against the operators of certain AI systems'[162]. However, as AI is growing and changing at an unprecedent pace, and as harms have already been occurring, it would be preferable to adopt now the most protective option (strict liability) and review the impact of the measure well before the five years.

The third critique questions whether the proposed AILD—with its limited improvements on the discovery and causal link fronts—adequately considers the current features of AI in the 'GenAI' season, as well as being a good fit for the harms that are most likely to occur. The proposed AILD is clearly designed to tackle, albeit imperfectly, the black box problem i.e. the intrinsic lack of transparency of AI algorithms that has given rise to such a large body of research that could be regarded as a field in its own right: explainable AI[163]. While opacity continues to be a problem for AI at the time of ChatGPT—and as such the easing of the burden of proof under the AILD would be a good response—our previous sections shone a light on a number of features and potential vulnerabilities that remain unaccounted for in the current liability system. Perhaps ironically for a technology that works by predicting the next word in a sentence, GenAI operates

[155] Product Liability Directive, art 7(f); and in the US context, Richard E Byrne, 'Strict Liability and the Scientifically Unknowable Risk' (1973) 57 Marquette Law Rev 660.

[156] For many years now, scholars have argued that strict liability would be the best response to AI harms. See e.g. Wendehorst (n 64); Herbert Zech, 'Liability for AI: Public Policy Considerations' (2021) 22 ERA Forum 147.

[157] AILD Explanatory Memorandum [1].

[158] Commission Staff Working Document Impact Assessment Report Accompanying the document Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (SWD/2022/319 final) [2.2].

[159] Hacker (n 152) 30. The author also argues for the limitation of strict liability to economic operators and professional users of high-risk AI systems. We are not convinced that linking the framework to the concept of high-risk systems is the best solution. As we have argued in this paper, risk is a useful tool to impose ex-ante duties, less so when it comes to harms that have already occurred.

[160] It has been argued that, while some advancement can appreciated on the fragmentation front, the proposed AILD does not bring about any meaningful improvement in terms of legal certainty (Marta Ziosi and others, 'The EU AI Liability Directive (AILD): Bridging Information Gaps' (2023) 14 Eur J Law Technol 1 <https://ejlt.org/index.php/ejlt/article/view/962> accessed 21 June 2024.).

[161] Hacker (n 152) 7. To support SMEs, the author suggests an exception to the proposed strict liability framework; namely, SMEs as well as operators and users of non-high-risk AI applications should only be covered by a presumption of defectiveness, breach of duty and causality' (ibid 33). Other exceptions to the strict liability rule would apply to actions against consumers using AI (for which fault-based liability is put forward) and foundation models, as we will see.

[162] Draft AILD, art 5(2).

[163] See e.g. Mayuri Mehta, Vasile Palade and Indranath Chatterjee (eds), *Explainable AI: Foundations, Methodologies and Applications* (Springer, Cham (CH) 2023).

in ways that often cannot be predicted by its own creators, as the aforementioned phenomena of emergence and zero-shot learning show. Predictability plays a key role when it comes to liability in tort. For example, the test of remoteness in English tort law has the function of identifying which consequences of the defendant's conduct the latter should shoulder, and unpredictable harms would be typically regarded as too remote to be eligible for compensation[164]. The rise of agentic AI—agents that thanks to LLMs are becoming increasingly akin to Artificial General Intelligence—are exacerbating the problem. In a world where artificial agents perform increasingly complex tasks on behalf of their users, and do so in a way that is close to fully autonomous, tortious liability must be rethought. Indeed, it is becoming difficult to understand how a human could (and should) be considered to be at fault for the agent's actions that are not known and controlled, let alone predicted. This corroborates our suggestion that only a strict liability system would be a suitable for the current stage of AI. Indeed, if the harm is unforeseeable, by its very nature there cannot be fault in the sense of a breach of a duty of care that was in place to prevent that harm from materializing. Strict liability would be better suited for agentic AI, and AI companies could still put forward the argument that the harm would fall under *force majeure* or fortuitous event. In conclusion, only strict liability would provide those incentives for AI companies to develop and deploy safe systems or, in other words, to take responsible AI seriously. The revision of the Product Liability Directive is already pointing in this direction: now it is a matter of extending the approach beyond the liability for harms caused by a defect in products.

A final argument in favour of strict liability for AI torts is that the law that regulates technology should learn from technology itself, as we noted when discussing the need for an incremental approach to regulation, one that mimics technology design. Similarly, autonomy and automation are at the core of AI, the legal framework should be as stripped back and code-friendly as possible. Whereas to ascertain fault is a complex and discretionary endeavour, strict liability is the simplest form of liability as it comes with less requirements and it is more straightforward. This is also another argument against the aforementioned proposal[165] of a gradation of ex-post liability based on risk: while rather elegant from a legal viewpoint, it does not lend itself for the simplicity that AI demands. As such, strict liability can provide near automatic redress in a way that mirrors the way AI works.

CONCLUSION

The lack of a general express competence to harmonize tort law[166] has not prevented the gradual emergence of EU tort law,[167] which is hardly surprising as similar phenomena have been observed in other areas, most notably Intellectual Property.[168] One of the reasons why the GDPR became the epitome of the Brussels effect was that it partly harmonized tort law in data-related scenarios.[169] Unless the EU corrects its course, it is unlikely that a Brussels effect will manifest also in the AI space.[170] If the AILD has the ambitious goal of 'adapt[ing] private law to the needs of the

[164] Andrew Tettenborn (ed), *Clerk & Lindsell on Torts* (24th edn, Sweet & Maxwell, London (UK) 2023) 2-140.
[165] Hacker (n 152).
[166] Bussani and Infantino (n 62) 5.
[167] Paula Giliker (ed), *Research handbook on EU tort law* (Edward Elgar, Cheltenham (UK) 2017); Gert Brüggemeier, *Tort law in the European Union* (2nd edn, Wolters Kluwer, Alphen aan den Rijn (NL) 2018).
[168] Ana Ramalho, 'Conceptualising the European Union's Competence in Copyright—What Can the EU Do?' (2014) 45 IIC 178.
[169] See Claudio Scognamiglio, 'Danno e Risarcimento Nel Sistema Del Rgpd: Un Primo Nucleo Di Disciplina Eurounitaria Della Responsabilita` Civile?' (2023) 5 NGCC 1150, describing the current state of things as a building site that will require additional interventions by the Court of Justice, national court, and academics (ibid 1159). His work focussed on case C-300/21 *UI v Österreichische Post AG* EU:C:2023:370, and the issue of compensability of non-pecuniary damage caused by the fear of potential misuse of personal data, but those observations have wider applicability.
[170] We would agree with the words of caution of Massimilano Granieri, 'Una Sinopsi Comparativa e Una Prospettiva Critica Sui Tentativi Di Regolazione Dell'intelligenza Artificiale' (2023) 2 Comp dir civ 703.

transition to the digital economy', it cannot succeed in its current form. We would agree with those commentators who have put forward that the EU approach to AI liability is half-hearted[171] and cumbersome,[172] and that the proposed AILD constitutes a 'a very small step forward […] a liability framework in the name only'[173]. However, the legislative process is ongoing and there still is the opportunity to rise to the challenge and proceed with a more systematic harmonization of tort law. By its very nature—i.e. an AI-related instrument—one may say that even a significant rewriting of the AILD to account for all the potential issues raised by AI-generated damages would not suffice as the law would remain fragmented every time an AI is not involved in the damage. To this, we would object that, as AI and IoT increasingly converge,[174] we will soon be living in a world where every object and service is equipped with some form of AI. We have observed similar phenomena in the past, most notably with cloud computing—initially a niche technology used just to save back-up copies of files, it now underpins the entire Internet infrastructure.[175] Harmonizing AI-related torts would effectively mean harmonizing tort law as a whole. If this is the potential impact, and learning from the lesson of the AI Act that was too rashly approved mainly for political reasons, we should pause, scrap the AILD, and start from scratch. Responsible AI demands a more thought-out and comprehensive approach to AI liability, one that learn from the way that technology development works i.e. by introducing innovations incrementally, test them, and learn from any errors and flaws. We believe that only fully harmonized strict liability would account for how AI works and for the risks it poses. If we do not go back to the drawing board and take the harmonization of torts seriously, the realization of AI safety and responsible AI will be under threat. Similarly, if general private law—including torts—does not take its regulatory role seriously, it risks being condemned to irrelevance. As AI becomes ubiquitous, lawmakers are under pressure to operate in a fast-faced and reactive way. We call on them to resist this temptation. We should apply Daniel Kahneman's slow thinking[176] to law-making, otherwise we may as well replace that too with an AI.

[171] Hacker (n 152).
[172] Kretschmer and others (n 12) 12.
[173] Li and Schütte (n 22) 151.
[174] See Guido Noto La Diega, 'IoT and AI in Privacy Law' in Ryan Abbott and Elizabeth Rothman (eds), *Elgar Concise Encyclopedia of Artificial Intelligence and the Law* (Edward Elgar 2025).
[175] One need only think of how ubiquitous (both in private and public ecosystems) Amazon's AWS has become, or that we now stream not only music and films, but also software itself (Software-as-a-Service or SaaS).
[176] See also Ngwako Ralepelle, 'Why Slow Thinking Matters for AI' (*Medium*, 2 February 2024) <https://medium.com/@nralepelle/why-slow-thinking-matters-for-ai-ba015c3bb84a> accessed 10 April 2024.