Stanford – Vienna
Transatlantic Technology Law Forum

A joint initiative of
Stanford Law School and the University of Vienna School of Law

European Union Law Working Papers

No. 76

The EU Regulatory Approach(es) to AI Liability

Maria Lillà Montagnani & Marie-Claire Najjar

2023

European Union Law Working Papers

Editors: Siegfried Fina and Roland Vogl

About the European Union Law Working Papers

The European Union Law Working Paper Series presents research on the law and policy of the European Union. The objective of the European Union Law Working Paper Series is to share "works in progress". The authors of the papers are solely responsible for the content of their contributions and may use the citation standards of their home country. The working papers can be found at http://ttlf.stanford.edu.

The European Union Law Working Paper Series is a joint initiative of Stanford Law School and the University of Vienna School of Law's LLM Program in European and International Business Law.

If you should have any questions regarding the European Union Law Working Paper Series, please contact Professor Dr. Siegfried Fina, Jean Monnet Professor of European Union Law, or Dr. Roland Vogl, Executive Director of the Stanford Program in Law, Science and Technology, at:

Stanford-Vienna Transatlantic Technology Law Forum
http://ttlf.stanford.edu

Stanford Law School
Crown Quadrangle
559 Nathan Abbott Way
Stanford, CA 94305-8610

University of Vienna School of Law
Department of Business Law
Schottenbastei 10-16
1010 Vienna, Austria

About the Authors

Maria Lillà Montagnani is Professor of Commercial Law at Bocconi University in Milan, where she teaches and does research in the field of Law and Technology, and acts as Director of the LL.M in European Business and Social Law. She is a Transatlantic Technology Law Forum Fellow at Stanford Law School. She has been a visiting professor at Peking University School of Transnational law and Haifa School of Law, Global Research Fellow at NYU Law School, faculty associate at the Berkman Klein Center of Harvard University and scholarship holder at the Max Planck Institute in Munich. Lillà has widely published in the field of IP and technology in national and international journals and prestigious book series; she is also on the editorial board of several law journals and acts as an independent expert with the European Commission (lastly in relation to AI liability).

Marie-Claire Najjar is Research assistant and Teaching assistant at Bocconi University in Milan, where she also acts as co-coordinator of the LL.M. in European Business and Social Law (of which she is an alumna). She is also specialising in Privacy, Cybersecurity and Data Management at Maastricht University (at the European Centre on Privacy and Cybersecurity). This builds a bridge between her multi-disciplinary studies in economics/computer science (B.Sc.) and business law (LL.M.) at Bocconi University, and her experience in product management of AI-powered products. Marie-Claire's fields of interest are Responsible AI, Emerging Digital Technologies and the law, and specifically the use of data analytics in sports (on which she is publishing a study in the Albany Law Journal of Science and Technology).

General Note about the Content

Suggested Citation

This European Union Law Working Paper should be cited as:
Maria Lillà Montagnani & Marie-Claire Najjar, The EU Regulatory Approach(es) to AI Liability, Stanford-Vienna European Union Law Working Paper No. 76, http://ttlf.stanford.edu.

Copyright

Abstract

The continued progress of Artificial Intelligence (AI) can benefit different aspects of society and various fields of the economy, yet pose crucial risks to both those who offer such technologies and those who use them. These risks are emphasized by the unpredictability of developments in AI technology (such as the increased level of autonomy of self-learning systems), which renders it even more difficult to build a comprehensive legal framework accounting for all potential legal and ethical issues arising from the use of AI. As such, enforcement authorities are facing increased difficulties in checking compliance with applicable legislation and assessing liability, due to the specific features of AI (and more generally, of emerging digital technologies) — namely: complexity, opacity, autonomy, predictability, openness, data-drivenness, and vulnerability.

This scenario challenges policymaking in an increasingly digital and global context, where it becomes difficult for regulators to predict and face the impact of AI systems on economy and society, to make sure that they are human-centric, ethical, explainable, sustainable and respectful of fundamental rights and values. The European Union has been dedicating increased attention to filling the gap between the existing legal framework and AI. Some of the legislative proposals in consideration call for preventive legislation and introduce obligations on different actors — such as the proposed AI Act and Machinery Regulation — while others have a compensatory scope and seek to build a liability framework — such as the proposed AI Liability Directive and revised Product Liability Directive.

The present paper starts by assessing the fit of the existing European liability regime with the constantly- evolving AI landscape, by identifying the normative foundations on which a liability regime for such technologies should be built on. It then addresses the proposed additions and revisions to the legislation, focusing on how they seek to govern AI systems. Finally, it considers potential additional measures that could continue to strike a balance between the interests of all parties, namely by seeking to reduce the inherent risks that accompany the use of AI and to leverage its major benefits for our society and economy.

Table of Contents

INTRODUCTION

In December 2020, an Italian labor union sued Deliveroo, a food delivery platform, over the use of an artificial intelligence system that assigned a "reliability rating" to its delivery workers.[1] The rating was used to determine the preference order in future shift scheduling, and the court found it discriminatory. It also held Deliveroo *liable* for the AI's output since it did not consider justifications provided under Italian labor law.

This case highlights the central role that artificial intelligence (AI)[2] plays in the gig economy[3] and the risks associated with its deployment when not properly supervised. AI systems develop in a way that makes them pursue their tasks with diverse degrees of autonomy.[4] Their new and enhanced potential brings in risks – or increase the existing ones – for both those who offer them and those who use them. Indeed, such technologies may have unintended effects or be used for malicious purposes. They can lead not only to discrimination and biases,[5] but also to violation of IP and personality rights,[6] unauthorized access and cybersecurity vulnerabilities,[7] and errors that can ruin a person's life.[8] As a matter of fact, while AI can bring

[1] The case was decided by the Bologna tribunal (sez. lavoro RG 2949/2019, ord. 31.12.2020) and it is available at: https://www.bollettinoadapt.it/wp-content/uploads/2021/01/Ordinanza-Bologna.pdf. For a comment *see* Pietrogiovanni, V. (2021) "Deliveroo and Riders' Strikes: Discriminations in the Age of Algorithms", *International Labor Rights Case Law*, 7(3), pp. 317-321. Available at: https://brill.com/view/journals/ilrc/7/3/article-p317_317.xml; Fava, G. (2021) "L'ordinanza del Tribunale di Bologna in merito alla possibile discriminatorietà dell'algoritmo utilizzato da Deliveroo", *Lavoro Diritti Europa*, 2021/1.

[2] According to the definition endorsed at European level, "[a]rtificial intelligence (AI) refers to systems that display intelligent behavior by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)." (High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and scientific disciplines* 1 (2019), available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341.).

[3] Ternullo, F. (2019). "AI Implications in GIG Economy (Uber Scenario and not only)" *European AI Alliance*, 8 February. Available at https://futurium.ec.europa.eu/en/european-ai-alliance/forum/ai-implications-gig-economy-uber-scenario-and-not-only?language=en; Liu, Y. *et al.* (2022) "Unintended consequences of Advances in Matching Technologies: Information Revelation and Strategic Participation on Gig-economy Platforms". Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3877868.

[4] On the concept of autonomy, see *infra* section I.A.

[5] Algorithms trained with biased data may lead to biased decisions to the detriment of minorities when screening job candidates, assessing creditworthiness for loans, or predicting criminal behavior. *See* Manyika, J., Silberg, J., and Presten, B. (2019) "What Do We Do About the Biases in AI?", *Harvard Business Review*. Available at: https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai; Even a chatbot may turn out to be racist. This occurred with Tay, the chatbot developed by Microsoft to self-learn conversational skills and autonomously interact with users via Twitter. It was shut down in 2016. *See* Lee, P. (2016) "Learning from Tay's introduction" [Blog] *Official Microsoft Blog*. Available at: https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/.

[6] Autocomplete functions of search engines may cause defamation, reputational damage, or trademark violations. *See* Karapapa, S., Borghi, M. (2015) "Search Engine Liability for Autocomplete Suggestions: Personality, Privacy and the Power of the Algorithm", *International Journal of Law and Information Technology*, 23(3), pp.261-289.

[7] For example, a flaw in the radio of a vehicle could expose the risk of unauthorised access by a third party maliciously intending to take over the control system of the self-driving car. For a similar case, *see* the RAPEX notification from Germany published in the EU Safety Gate website (A12/1671/15) (https://ec.europa.eu/safety-gate-alerts/screen/webReport/alertDetail/188127). Similarly, cyberattacks on the control systems of driverless metro, autonomous weapons, industrial plants, or critical infrastructures may cause enormous damage as well, if not properly governed.

[8] On the possibility that a robo-advisor leads to wrong investments see Litz,D. (2018) "Risk, Reward, Robo-Advisers: Are Automated Investment Platforms Acting in Your Best Interest", *Journal of High Technology Law*, 18(2), p. 367. Also, errors in

benefits such as increased productivity and can transform products, services, activities, procedures, and practices in several economic sectors and in relation to many aspects of society – such as health,[9] sustainability,[10] sports,[11] transportation [12], and the like, it can also lead to unintended effects, biases, and violations of fundamental rights.

The double-edged nature of AI raises challenges for regulators and policymakers who need to balance its potential benefits with its potential harms. This requires a "responsible innovation"[13] strategy that ensures AI is human-centric, ethical, explainable, sustainable, and respectful of fundamental rights and values.[14] To create an environment of trust and accountability, legal rules on civil liability must be designed to address the risks generated by AI-based technologies.

To this end, the European Union (EU) has already adopted initiatives under the AI strategy to foster the development of AI while addressing the impact on fundamental rights.[15] The Commission presented its AI

automated diagnoses and surgeries may ruin a person's life. Autonomous vehicles, although promising a reduction of accidents caused by human errors, could still cause accidents due to flaws in object recognition technologies embedded in self-driving.

[9] AI can assist in diagnoses, for instance in oncology. Some AI detection systems – namely one for melanoma diagnosis – have been found to perform even better than humans. See "Artificial intelligence will improve medical treatments", *The Economist*, 9 June 2018. Available at: https://www.economist.com/science-and-technology/2018/06/09/artificial-intelligence-will-improve-medical-treatments?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=18151738051&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gclid=CjwKCAiA5sieBhBnEiwAR9oh2k25zTTED3lIEzVZjkPYK7vphdS9klvkbD5r-WbWE78ED0ZMizCCIhoCp20QAvD_BwE&gclsrc=aw.ds. Similarly, the Internet of Bodies – that is, the merger of IoT and AI with the human body – can allow patients to be reminded if they forgot to ingest their medication on the basis of a sensor implanted in their stomach. *See* Matwyshyn, A.M. (2019) "The Internet of Bodies", *William & Mary Law Review*, 61(1), p. 81

[10] AI can teach itself to reduce the use of energy, required (for instance) by data centers. This is the case of the AI developed by Google's DeepMind division. See Gow, G. (2020) "Environmental Sustainability And AI", *Forbes*, 21 August. Available at: https://www.forbes.com/sites/glenngow/2020/08/21/environmental-sustainability-and-ai/?sh=26938a717db3.

[11] The Internet of Bodies can allow athletes to track their performance with wearable devices. See Fierens, M., De Bruyne, J. (2021) "Legal and Ethical Considerations Concerning AI in Sports", *Sports Tech Research Network,* 21 December. Available at: https://sports-tech-research-network.com/news-insights/2020/12/21/Legal-and-ethical-considerations-concerning-AI-in-sports.

[12] For example, the Internet of Bodies can allow truck companies to check the alertness of their drivers. *See* Matwyshyn, *supra* note 9, at p. 84.

[13] This concept, intended to emphasize the role of responsibility in shaping and promoting innovation, is gaining increasing attention among scholars and policy makers. On this, *see* Koops, B.J., *et al.* (eds) (2015) *Responsible Innovation 2: Concepts, approaches, and Applications*. Dordrecht: Springer.

[14] There is a lively discourse around ethical issues raised by new technologies, as analyzed in the UNESCO's *Recommendation on the Ethics of Artificial Intelligence* (2021). Available at https://unesdoc.unesco.org/ark:/48223/pf0000380455. *See* also Dubber M.D., Pasquale F., Das S. (Eds) (2020) *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press; and Coeckelbergh, M. (2020) *AI Ethics*. Cambridge, MA: The MIT Press. With specific regard to algorithmic transparency and the need to shift from a black-box society to an intelligent one, *see* Pasquale, F. (2015) *The Black Box Society*. Cambridge, MA: Harvard University Press, 218 ("Rather than contort ourselves to fit 'an impersonal economy lacking a truly human purpose,' we might ask how institutions could be re- shaped to meet higher ends than shareholder value . . . Black box services are often wondrous to behold, but our black box society has become dangerously unstable, unfair, and unproductive. Neither New York quants nor California engineers can deliver a sound economy or a secure society. Those are the tasks of a citizenry, which can perform its job only as well as it understands the stakes").

[15] For an overview of the European AI strategy *see* https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence. For a first comment *see* Justo-Hanani, R. (2022) "The politics of Artificial Intelligence regulation and governance reform in the European Union", *Policy Sciences*, 55, pp.137–159. Available at: https://doi.org/10.1007/s11077-022-09452-8.

package in April 2021, which includes the proposal for an AI Regulation (also "AI Act" or "Regulation").[16] In addition, the EU has recently complemented the AI package with two further legislative instruments: a proposal for a revised Product Liability Directive ("revised PLD")[17] and a proposal for an AI Liability Directive ("AILD").[18] The several measures mentioned contribute to the adoption of an AI liability regime at EU level, with the intent of establishing a framework that by avoiding under- or over-compensation of victims, can achieve an environment of trust necessary for a development of AI that is beneficial to the economy and the society.

In the following sections, we first discuss the main features of AI and the challenges they represent for traditional liability notions (section I). We then illustrate the development of the EU's approach to AI liability, from the established liability framework (section II) to the new proposals (section III). Finally, we conclude by providing an assessment of the proposed AI liability framework.

I. New Technological Features vis-à-vis Traditional Notions and Regime(s) of Liability

In this section we first analyse how AI features challenge the traditional notions of liability and second the state of the art as to the current liability regime(s) in Europe.[19] We do this to set the ground for understanding the need to adopt new – and harmonized – rules.

[16] This initial step (in April 2021) encompasses the following documents: Communication on fostering a European approach to Artificial Intelligence, COM/2021/205 final (Communication on fostering a European approach to AI). Available at: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2021%3A205%3AFIN; Annexes to the Communication, including a 2021 Review of the Coordinated Plan on Artificial Intelligence. Available at: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review (2021 Review of the Coordinated Plan on AI); Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, COM (2021)206. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (AI Act or Regulation); and Impact Assessment of the Regulation on Artificial intelligence – Commission Staff Working Document. Available at: https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence (Impact Assessment accompanying the AI Act).

[17] Proposal for a Directive of the European Parliament and of the Council on liability for defective products, COM/2022/495 final, on 28 September 2022. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0495 (Revised PLD).

[18] Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM/2022/496 final, on 28 September 2022. Available at: https://commission.europa.eu/system/files/2022-09/1_1_197605_prop_dir_ai_en.pdf (AILD).

[19] We use the plural "regimes" as this is not an harmonized area of law at EU level. *See*, *ex multis*, Cafaggi, F., (2010) "Private Regulation in European Private Law" in Hartkamp, A. *et al.* (eds) (2011), *Towards a European Civil Code*, Available at: https://ssrn.com/abstract=1443948.

A. Old Dogs, New Tricks – Old Notions, New Features

The question as to whether current liability regime(s) in Europe are fit for the new digital era comes from the fact that AI presents features that are unknown to the previous generation of technologies. Indeed, the main features of AI – namely complexity, opacity, autonomy, unpredictability, openness, data-drivenness, and vulnerability – challenge traditional liability notions such as damages, causal link, and duty of care.

In particular, AI is, in the first place, data-driven.[20] First, regarding the stage of training the model, issues can arise in data management and preparation, notably when the data used for the model is inaccurate, non-representative and insufficient, mirroring biases present in society, or unsecured and unprotected.[21] Inaccuracy, for instance, is often the product of inaccurate labelling, i.e. "the process by which the training data is manually assigned class labels", as we would end up with a "skewed ground truth" as a starting point.[22] Furthermore, to operate and self-develop, AI depends on information that is not pre-installed but generated by external or internal sources (like built-in sensors). This leads to issues whenever data is flawed or missing, due to an error in communication or in relation to the source. Moreover, AI is not completed once put into circulation. It depends upon subsequent inputs, such as updates and upgrades, and thus need to interact with other systems or data sources in order to operate. Its openness "by design" permits external input via some hardware plugin or wireless connection.[23] This constant interaction with outside information is what also makes these new technologies vulnerable to cybersecurity breaches, which can cause the systems to malfunction and/or modify its features in a way likely to cause harm.[24]

The features of data drivenness, vulnerability and openness challenge the traditional notion of damage – such as harm to persons and properties – as they enable the harm of further categories of protected interests, such as privacy, confidential information, security and cybersecurity. Indeed, there are built-in

[20] Sambasivan, N. *et al.* (2021) "Everyone wants to do the model work, not the data work: data cascades in high-stakes AI", *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, May 2021, Association for Computing Machinery, pp. 1–15. Available at: 10.1145/3411764.3445518; Ameen, S., *et al.* (2022) "AI Diagnostic Technologies and the Gap in Colorectal Cancer Screening Participation", in *Challenges of Trustable AI and Added-Value on Health*. Amsterdam: IOS Press, pp. 803-804, maintaining that "[d]ata is the critical infrastructure necessary to build AI systems", with issues arising from "AI/ML practices that undervalue data quality."

[21] Jacobs, M. and Simon, J. (2022) "Assigning Obligations in AI Regulation: A Discussion of Two Frameworks Proposed by the European Commission" *Digital Society* 1, 6. Available at: https://doi.org/10.1007/s44206-022-00009-z.

[22] Barocas, S., and Selbst, A. D. (2016) "Big data's disparate impact", *California Law Review* 104, pp.671-732.

[23] These open systems come as hybrid combinations of hardware, software, ongoing updates and services, as explained in Geistfeld, M. *et al.* (eds.) (2023), *Civil Liability for Artificial Intelligence and Software*. Berlin, Boston: De Gruyter, pp. 549-550.

[24] Lohn, A. (2020) "Hacking AI: A Primer for Policymakers on Machine Learning Cybersecurity", *Center for Security and Emerging Technology*. Available at: https://doi.org/10.51593/2020CA006.

features allowing systems to access and further interact with external information (for instance through updates). These features increase the systems' vulnerability to cybersecurity breaches, possibly leading to malfunction or feature modification likely to cause damage.[25] Similarly, cybersecurity breaches could imply the leaking of data, including personal or confidential data, which is further exacerbated by the systems' data-drivenness.

In the second place, whenever AI presents a self-learning capacity, it develops the ability to interpret the environment, interact with humans, cooperate with other actors, learn new behaviours, and execute actions without – or with limited - human intervention.[26] In this case AI becomes autonomous,[27] which in turn makes it unpredictable in its behaviour.[28] Indeed, many systems are designed to identify and classify new stimuli and link them to self-chosen – not pre-programmed – reactions. They rely on training data and data collected while interacting with surrounding environments, which in turn alters the initial algorithms. As a result, the more external data systems can process, the more unpredictable they become. Moreover, AIs presenting these features tend to be opaque as to their functioning due to the black box nature that they develop.[29] Opacity of AI systems may only increase in the presence of self-learning features, as algorithms no longer come as readable code but amount to black boxes that are almost impossible to understand. In addition, AI systems can also present an high degree of complexity whenever there is interdependency

[25] Geistfeld, M. *et al.* (2023), *supra* note 23, at 551.

[26] Most of the current applications of AI are based on supervised machine learning: the AI is trained on labelled data sets, on a specific topic, to learn rules and exceptions. Instead, when AI is based on unsupervised ML, it learns for itself from unlabeled data, i.e. from the data environment that it is placed into it sees new patterns in information that were not previously visible nor pre-defined. Moreover, reinforcement learning presents even more limited supervision that does not rely on training data. It instead works toward a goal through trial and error until it is consistently receiving a reward. *See* Smith, C.S. (2020) "Computers already learn from us. But can they teach themselves?", *The New York Times*, 8 April. Available at: https://www.nytimes.com/2020/04/08/technology/ai-computers-learning-supervised-unsupervised.html. Examples of Self-Learning AI Models can be found, for instance, in the medical field: Matheson, R. (2018) "Artificial intelligence model 'learns' from patient data to make cancer treatment less toxic", *MIT News,* 9 August. Available at: https://news.mit.edu/2018/artificial-intelligence-model-learns-patient-data-cancer-treatment-less-toxic-0810.

[27] For the development of autonomy in the medical sector *see* Bitterman D., Aerts H., Mak R. (2020) "Approaching autonomy in medical artificial intelligence", *The Lancet Digital Health* 2(9), pp. 447-449. Available at: https://doi.org/10.1016/S2589-7500(20)30187-4.

[28] Autonomous capabilities and intelligence ungoverned by human directions or supervision could lead to unexpected outcomes, as shown by the story of Alice and Bob, i.e., two chatbots developed to learn autonomous bargaining skills that started to interact using their own code, indecipherable for humans. *See* Griffin, A. (2017) 'Facebook's artificial intelligence robots shut down after they start talking to each other in their own language", *The Independent*, 31 July. Available at: https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research- openai-google-a7869706.html.

[29] Pasquale (2015) *supra,* note 14. More recently, Vaassen, B. (2022) "AI, Opacity, and Personal Autonomy", *Philosophy & Technology*, 35, 88. Available at: https://doi.org/10.1007/s13347-022-00577-5.

between different components and layers.[30] This increases the variety of players involved and complicates the understanding of potentially harmful processes.

The features of autonomy, unpredictability, opacity and complexity challenge the notions of causation and duty of care.[31] As to the former, liability regimes pivot around the principle that the victim should prove that the damage originated by some conduct or risk is attributable to the defendant. However, providing evidence of causation can become difficult with interconnected devices like AVs (combining hardware, software, connectivity and data) or self-learning AI systems (fuelled by machine learning – including deep learning – techniques and based on multiple external data collection). AI-empowered products may act in ways that were not envisaged at the time that the system was first put into operation, and these behaviours may be so autonomous to interrupt the causal link. The combination of opacity and complexity leads to issues in establishing causation, due to the variety of actors and of causes (possibly successive) which could have contributed to the damage.[32]

As to the latter, the duty of care is central in fault-based liability regime and requires the adherence to a standard of reasonable care while performing any acts that could foreseeably harm others.[33] While statutory language may in certain cases define such duties, in many others they are reconstructed by the court based on social beliefs about the prudent and reasonable course of action in the circumstances at stake. Applying fault-based liability rules to AI systems is difficult, because they lack well-established models of proper functioning and develop by learning without direct human control. The processes running them cannot all be measured according to duties of care designed for human conduct, an accepted standard of care for the creation and operation of autonomous systems has not emerged yet. It may sometimes be hard even to identify the person obliged to meet such duty of care. In fact, it could be unfair or inefficient to assign liability for any damage caused by an AI product always to the designer of the algorithm. Depending on

[30] These components can range from tangible parts and devices (e.g. sensors, actuators, hardware), to software components, data, and connectivity features. This is described, for EDTs in general, in the Commission Staff Working Document, *Liability for emerging digital technologies*, accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial intelligence for Europe, SWD/2018/137 final (2018). Available at: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A52018SC0137.

[31] Geistfeld, M. *et al.* (2023), *supra* note 23.

[32] Koch B.A., *et al.* (2022) "Response of the European Law Institute to the Public Consultation on Civil Liability – Adapting Liability Rules to the Digital Age and Artificial Intelligence", *Journal of European Tort Law* 13(1), pp. 25-63. Available at: https://doi.org/10.1515/jetl-2022-0002.

[33] Geistfeld, M. *et al.* (2023), *supra* note 23.

circumstances liability should be allocated also to owners and/or users, yet this is not self-evident with features of AI systems.[34] However, according to any liability regime, tracing a damage back to a specific person is still a fundamental prerequisite for any fault-based claim.[35]

All the above-mentioned features make it more difficult for market surveillance and enforcement authorities to assess liability and compliance with applicable legislation unless they are adequately governed. A first answer to the challenges above illustrated comes from the scholars that suggested to update the traditional liability notions to align them to the technological pace. In particular, one school of thought pushes for joint and several liability of all subjects involved in the design, programming and deployment of an AI application.[36] On one hand, this would facilitate claims of compensation for damage; on the other, it might be ineffective in allocating costs and setting prevention incentives for all relevant players.[37] Some scholars, instead, urge to reconceptualize intelligent and autonomous machines as entities with the status of a "person" under the law, such that AI can be held directly liable for harm – just as legal entities are.[38] They argue that an intelligence "even able to supersede humans in a number of areas" could sometimes be at fault.

However, this legal fiction may open up more problems than it solves, particularly as to the definition of selection criteria and equity requirements, as well as to the allocation of costs among all parties involved in

[34] For instance, with regard to autonomous weapons, "somehow human responsibility and accountability for the actions taken by the machine evaporate and disappear. The soldier in the field cannot be expected to understand in any serious way the programming of the machine; the designers and programmers operate on a completely different legal standard; the operational planners could not know exactly how the machine would perform in the fog of war; and finally, there might be no human actors left standing to hold accountable" (Anderson, K., Waxman, M.C. (2017) *"Debating Autonomous Weapon Systems, their Ethics, and their Regulation under International Law",* in Brownsword R., Scotford, E., Yeung, K. (eds.), *The Oxford Handbook of Law, Regulation and Technology*. Oxford: Oxford University Press, pp. 1097-1110).

[35] European Commission, *On Artificial Intelligence - A European approach to excellence and trust*, 19.02.20, COM (2020) 65 final. Available at: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

[36] Vladeck, D.C. (2014) "Machines Without Principals: Liability Rules and Artificial Intelligence", *Washington Law Review* 89, pp. 117-149.

[37] Comandé, G. (2019) "Multilayered (Accountable) Liability for Artificial Intelligence*"*, in Lohsse, S., Schulze, R., and Staudenmayer, D. (eds.), *Liability for Artificial Intelligence and the Internet of Things*. Oxford: Hart Publishing; Baden-Baden: Nomos, pp. 165-175.

[38] *See generally*, Hage, J. (2017) "Theoretical foundations for the responsibility of autonomous agents*"*, *Artificial Intelligence and Law* 25, pp. 255-271; Jackson, B.W. (2019) "Artificial Intelligence and the Fog of Innovation: A Deep-Dive on Governance and the Liability of Autonomous Systems", *Santa Clara High Technology Law Journal* 35(4), pp. 35-63. This debate has a long history, as shown by Solum, L.B. (1992) "Legal Personhood for Artificial Intelligences", *North Carolina Law Review* 70(4), pp. 1231-1287. A case against treating robots like humans is made by Eidenmuller, H. (2017) "The Rise of Robots and the Law of Humans", Oxford Legal Studies Research Paper (27).

the development and use of AI applications.[39] So far, legislators and courts seem far from revolutionizing the traditional notions of liability to introduce some sort of robot's fault.

There have been proposals to apply a "reasonable algorithm" standard to self-learning systems, given their similarity to humans in decision-making and the consequent damage. However, this poses a crucial, so far unresolved question: what could be considered reasonable behavior for algorithms?[40] Rather than resorting to conceptually new theories, another – maybe more viable – option that has been proposed is that of introducing a predetermined, detailed and acceptable level of care (or quasi-safe-harbor) for designers, manufacturers, owners and users of AI. If the level of care is unmet, a presumption of negligence and, therefore, liability would be triggered; if met, the defendant would enjoy a quasi-safe harbor, while the claimant would bear the burden of proving actual negligence.[41]

B. Traditional Liability Regime(s)

The notions of damage, causality and duty of care above illustrated are rooted in the national liability regimes of Member States as at EU level a liability framework is only partially harmonised.[42] Indeed, the existing EU tort law rules are currently limited – at least until the approval of the proposed framework on AI liability [43] – to the current version of product liability under Directive 85/374/EC ("PLD")[44], liability

[39] Comandè, G. (2019) "Intelligenza artificiale e responsabilità: Il carattere trasformativo dell'IA e il problema della responsabilità", *Analisi giuridica dell'economia* 1, pp. 169-179.
[40] Chagal-Feferkorn, K.A. (2021) "How Can I Tell if My Algorithm Was Reasonable?", *Michigan Technology Law Review* 27, pp. 213-261. Available at: https://repository.law.umich.edu/mtlr/vol27/iss2/2.
[41] Rachum-Twaig, O. (2020) "Whose Robot Is It Anyway?: Liability for Artificial-Intelligence-Based Robots*"*, *University of Illinois Law Review*, 11(7), p. 1141, in particular pp.1172-73.
[42] Oliphant, K. (2012) "Cultures of Tort Law in Europe", *Journal of European Tort Law*, 3 pp.147-157; Bussani, M. and Infantino, M. (2015) "Tort Law and Legal Cultures", *American Journal of Comparative Law*, 63(1), pp. 77-108.
[43] See *infra* section III.
[44] 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 O.J. (L 210) ("PLD"). A thorough account of such instrument is made in Fairgrieve, D., *et al.* (2016) "Product Liability Directive", in Machnikowski, P. (ed.), *European Product Liability: An Analysis of the State of the Art in the Era of New Technologies*. Cambridge: Intersentia, pp. 17-108. *See also* Faure, M.G. "*Economic Analysis of Product Liability"*, in Machnikowski, P. (ed.) *European Product Liability: An Analysis of the State of the Art in the Era of New Technologies*. Cambridge: Intersentia, pp. 619-666.

for infringing data protection law (under Article 82 of the GDPR),[45] and liability for infringing competition law (under Directive 2014/104/EU).[46]

As a result, harmful damages arising during the use of AI are likely to be compensated under existing national tort and contract law, or, where applicable, through the liability provisions of Member States. For example, some national jurisdictions have specifically regulated the use of AVs, also providing for coverage of any damages caused, by insurance or by reference to the general rules.[47]

In general, domestic tort laws include a rule introducing fault-based liability with a broad scope of application, accompanied by several more specific rules which either modify the premises of fault-liability (especially in the distribution of the burden of proof) or establish liability independently from fault (strict or risk-based liability). Most liability regimes also encompass the notion of liability for others (indirect or vicarious liability), which can, in turn, be fault- or risk-based, depending on the case or the country.

It is worth pointing out that all Member States' liability frameworks share some common principles. A general rule of fault-based liability is a common ground.[48] When an actor fails to take due care, and this negligence causes harm to another – or they cause such harm intentionally – this actor is liable to compensate the victim. Usually, what triggers liability is harm to the fundamental interests of a person, such as life, health, bodily integrity, freedom of movement, private property, and in some countries also purely

[45] Commission Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) (GDPR).

[46] Directive 2014/104/EU of the European Parliament and of the Council of 26 November 2014 on Certain Rules Governing Actions for Damages under National Law for Infringements of the Competition Law Provisions of the Member States and of the European Union Text with EEA relevance, 2014 O.J. (L 349).

[47] *See* for example the Germany amended its Road Traffic Act to allow driverless vehicles on public roads: Act Amending the Road Traffic Act and the Compulsory Insurance Act (Autonomous Driving Act), July 2021. Bundesgesetzblatt Jahrgang 2021 Teil I Nr. 48, ausgegeben zu Bonn am 27. Juli 2021. Available at: https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl121s3108.pdf#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl121s3108.pdf%27%5D__1676679941289. In France, too, the framework for the deployment of automated vehicles was initiated through a decree: Décret n° 2021-873 du 29 juin 2021 portant application de l'ordonnance n° 2021-443 du 14 avril 2021 relative au régime de responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite et à ses conditions d'utilisation. Available at: https ://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043729532. Also the UK announced plans to update the Highway Code, in order to encompass drivers' responsibilities in self-driving vehicles. See Department for Transport, Centre for Connected and Autonomous Vehicles, and Trudy Harrison MP (2022), "Britain moves closer to a self-driving revolution", *Gov.uk*, 20 April. Available at: https://www.gov.uk/government/news/britain-moves-closer-to-a-self-driving-revolution. *See*, among many, Diehl, R., and Thue, M.I. (2017) "Autonomous Vehicle Testing Legislation: A Review of Best Practices from States on the Cutting Edge", *Journal of Technology Law & Policy*, 21(2), pp.197-221; and, in the US: Geistfeld, M.A. (2017) "A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation", *California Law Review* 105, pp.1611-1694.

[48] European Group on Tort Law (ed.) Principles of European Tort Law, Art. 1.101 (1)-(2). For a comment, *see* Busnelli, F.D. *et al.* (eds.) (2005) *Principles of European Tort Law: Text and Commentary*. Vienna: Springer.

economic losses and harm to human dignity. However, negligence and fault can be given different interpretations across Member States.

In addition, all Member States' legal systems encompass product liability as a result of the current PLD implementation which however dates back to 1984. On this base, a damage claim for harm generated by a defective product does not require a finding of fault on the part of the manufacturer, as, in principle, the PLD should introduce a harmonized strict – not fault-based – liability regime for defective products.[49] However, the regime that the PLD introduces resembles more a watered-down version of negligence liability than a strict liability regime since a claimant must, in any case, prove the defect and that such defect generates the harm that she suffered.[50] In sum, the current product liability regime only covers damages generated by defective products, leaving outside the provision of services, for which then the default negligence-based regime revives. Moreover, the PLD implementation not only has not been consistent in all Member States, but it lacks to cover instances generated by the use of AI, [51] and, more in general, all emerging digital technologies (EDTs).[52]

As a result, the EU scenario in force before the last wave of proposals is quite fragmented and in need of revision. Disparities in Member States' legislation and case-law concerning liability (fault-based, strict and vicarious) may produce distortions of competition and impair the functioning of the single digital market, while the moderate pace of European legislative harmonization may no longer be suitable to the rapid changes brought by AI.[53]

[49] PLD, *supra* note 44, at Recital 2, "liability without fault".
[50] PLD, *supra* note 44, at art. 4.
[51] For a survey of the issues as to the application of the Product Liability Directive to the EDTs, *see* De Meeus, C. (2019). "The Product Liability Directive at the Age of the Digital Industrial Revolution: Fit for Innovation?", *Journal of European Consumer and Market Law*, 8(4), pp.149-154.
[52] The category of emerging digital technologies is not fully defined and exhaustively identified in European documents on the topic, where they are indicated with the exemplificative list of "Internet of Things (IoT), Artificial Intelligence, advanced robotics and autonomous systems". In this work, the wording of the EU institutions is adopted.
[53] Morais Carvalho, J., and Nemeth, K. (2019). "Time for a Change? Product Liability in the Digital Era", *Journal of European Consumer and Market Law*, 8(4), pp. 160-161.

II. The Complementarity between AI Governance and the AI Liability Regime in the EU Policies

The need to address the issue of AI liability within the *acquis communitaire* – which initially emerged at policy level – has led to the adoption of three main proposals: the AI Act[54] – coupled with the revision of Machinery Regulation[55] – the revised PLD[56] and the directive on AI liability,[57] the former as a means of ex ante regulation, the latter two as means of ex post regulation. In fact, while the AI Act proposes provisions to govern AI, particularly high-risk AI systems, the directives propose rules to face the scenario in which a lack of AI governance generated damages.

In this section we illustrate the evolution of the EU policies as they constitute the background to a comprehensive AI liability framework. Indeed, these policies stress the importance for the regulatory framework to account for the legal issues raised by AI (including questions of liability).

A. Towards an Environment of Trust for AI

The debate on whether the current liability regime is fit for accommodating the issues previously described has been quite lively within the EU,[58] in particular as to what extent the existing liability schemes are adapted to the emerging market realities that follow the development of EDTs in general. In February 2017, the Resolution on Civil Law Rules on Robotics with recommendation to the Commission[59] proposed a whole range of legislative and non-legislative initiatives in the field of robotics and AI. A year later, in February 2018, the European Parliamentary Research Service study on a common EU approach to liability rules and insurance for connected and autonomous vehicles[60] was adopted as an added value assessment

[54] AI Act, *supra* note 16.
[55] Proposal for a Regulation of the European Parliament and of the Council on machinery products, COM/2021/202 final, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0202
[56] Revised PLD, *supra* note 17.
[57] AILD, *supra* note 18.
[58] For an analysis of the extent to which tort law may provide remedies to subjects injured by new technologies in the common law (Anglo-American) tradition *see* Morgan, J. (2017) "Torts and Technology", in Brownsword R., Scotford, E., Yeung, K. (eds.), *The Oxford Handbook of Law, Regulation and Technology*. Oxford: Oxford University Press, pp. 522-545.
[59] Resolution on Civil Law Rules on Robotics, EUR. PARL. DOC. 2015/2103(INL) (2017), http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html.
[60] Evas, T. (2018) "A common EU approach to liability rules and insurance for connected and autonomous vehicles", *European Added Value Assessment Accompanying the European Parliament's Legislative Own-initiative Report, European Parliamentary Research Service*, 2018, pp. 615-635. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2018/615635/EPRS_STU(2018)615635_EN.pdf.

accompanying the Resolution on Civil Law Rules. On April 25 2018, the Commission published a Staff Working Document on Liability for Emerging Digital Technologies,[61] accompanying the Commission's Communication on Artificial Intelligence for Europe,[62] which provides the starting point of the discussions on liability in the context initially of EDTs in general and later on specifically on AI.

All these documents, as well as the following Sibiu Communication of May 2019,[63] stress that a robust regulatory framework should address the ethical and legal questions surrounding AI, including those related to liability. In its 2018 AI Communication, the Commission also announced the adoption of a report assessing the implications of emerging digital technologies on existing safety and liability frameworks by mid-2019. In its 2019 Work Programme, it confirmed it would "continue work on the emerging challenge of Artificial Intelligence by enabling coordinated action across the European Union."[64] Accordingly, on April 2019, the high-level Expert Group on Artificial Intelligence set up by the European Commission listed liability frameworks among the non-technical methods for securing and maintaining a lawful and trustworthy AI,[65] on the assumption that an environment of trust is crucial for fully reaping the benefits of innovation.[66]

In order to provide an answer on how the liability regime could assist in achieving an environment of trust, in March 2018, the Commission also set up an Expert Group on Liability and New Technologies, operating in two different formations: the Product Liability Directive formation and the New Technologies formation.[67] This second formation was in particular asked to assess "whether and to what extent existing liability schemes are adapted to the emerging market realities following the development of the new

[61] Commission Staff Working Document, *Liability for emerging digital technologies*, accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial intelligence for Europe, SWD (2018) 137 final (Apr. 25 2018). Available at: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52018SC0137.
[62] Communication of the European Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial intelligence for Europe, COM (2018) 237 final (Apr. 25, 2018). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN.
[63] European Commission, *Preparing for a more united, stronger and more democratic Union in an increasingly uncertain world, Contribution to the informal EU27 leaders' meeting in Sibiu (Romania),* (May 9, 2019). Available at: https://ec.europa.eu/commission/sites/beta-political/files/euco_sibiu_communication_en.pdf.
[64] Communication of the European Commission, *Commission Work Programme 2019: delivering what we promised and preparing for the future*, COM (2018) 800 final, (Oct. 23, 2018). Available at: https://ec.europa.eu/info/sites/info/files/cwp_2019_en.pdf.
[65] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, at 6, 22, (Apr. 8, 2019). Available at: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines.
[66] Communication of the European Commission, *Building Trust in Human-Centric Artificial Intelligence*, COM (2019) 168 final (Apr. 8, 2019). Available at: https://ec.europa.eu/jrc/communities/sites/jrccties/files/ec_ai_ethics_communication_8_april_2019.pdf.
[67] *See* European Commission Expert Groups, *Expert Group on liability and new technologies*, (Mar. 9, 2018). Available at: https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3592.

technologies such as Artificial Intelligence, advanced robotics, the IoT and cybersecurity issues."[68] The experts were requested to examine whether the current liability regimes are still "adequate to facilitate the uptake of … new technologies by fostering investment stability and users' trust."[69] In case of shortcomings, the expert group was invited to make recommendations for amendments, without being limited to existing national and EU legal instruments. However, recommendations were to be limited to matters of extracontractual liability, leaving aside in particular corresponding (and complementary) rules on safety and other technical standards. As a result of the expert group's activity, in November 2019 the Report "Liability for Artificial Intelligence and other Emerging Digital Technologies" was published.[70] This undertook an assessment of existing liability regimes in the wake of emerging technologies and it concluded that the current ones in force in the Member States ensured at least basic protection of victims whose damage is caused by the operation of such new technologies, while also hinting to some adjustments that might be needed.

B. A European Approach to Artificial Intelligence

The need for some adjustments was confirmed also in the White Paper on artificial intelligence to foster excellence and trust[71] and in the Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics.[72] Both documents stress that the ultimate goal is to ensure remediation of damage caused by AI and overall reliability, while promoting investment stability and, more generally, innovation. In this context, efficient liability rules are deemed paramount for trustworthiness, which in turn is a prerequisite for the uptake of AI. Pursuing such a strategy was also defined a crucial step to strengthen

[68] *See* European Commission Expert Groups, *Call for Applications for the Selection of Members of the Expert Group on Liability and New Technologies*, (E03592) (Mar. 9, 2018). Available at: https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3592

[69] *Id.*

[70] Report of the Expert Group on Liability and New Technologies - New Technologies Formation, *Liability for Artificial Intelligence and other Emerging Digital Technologies* (2019). Available at: https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608.

[71] White Paper, *supra* note 35.

[72] Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee, *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, COM (2020) 64 final (Feb. 19, 2020). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2020%3A64%3AFIN.

European technology sovereignty and affirms the role of the EU on the international stage as "the most attractive, secure and dynamic data-agile economy in the world."[73]

For the purpose of achieving these goals, the European Commission suggested a regulatory and investment-oriented approach, entailing, among other things, adjustments to current European and national liability regimes. Indeed, a fragmented legal landscape sprinkled of different national initiatives could lead to the fragmentation of the single market and, consequently, endanger not just legal certainty, but also the emergence of a dynamic and flourishing European industry. Hence, the European Commission stressed the importance of aligning the efforts at European, national, and regional level,[74] while promoting partnership between the private and the public sector towards an "ecosystem of excellence" with proper incentives to research, innovation and deployment,[75] an "ecosystem of trust" duly protecting fundamental rights and consumers' rights[76] such as privacy and non-discrimination,[77] and through liability rules.

In line with the Report from the expert group, the European Commission's analysis of the current legal frameworks concluded for the adaptations of current norms and the adoption of new specific legislation, pursuing a targeted, risk- based approach, and ensuring effective enforcement. In order to address both current and anticipated technological, societal and commercial developments, such revised regulatory

---

[73] White Paper, *supra* note 35, at p. 3.

[74] Stronger coordination is encouraged in Communication of the European Commission, *see* Coordinated Plan on Artificial Intelligence, COM (2018) 795 final (Dec. 7, 2018), available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56018, regarding a plan among the European Commission, Member States, Norway, and Switzerland for some 70 joint actions in the following key areas: (i) increasing investment, (ii) making more data available, (iii) fostering talent, and (iv) ensuring trust. The plan will run until 2027, with regular monitoring and update.

[75] To foster investments, the European Commission has proposed a number of measures under the Digital Europe Programme, Horizon Europe and the Multiannual Financial Framework for 2021 to 2027. On this, *see* European Commission, *Info session Horizon 2020: Artificial intelligence for manufacturing*, (Nov. 18, 2019). Available at: https://ec.europa.eu/digital-single-market/en/news/info-session-horizon-2020-artificial-intelligence-manufacturing. A key role is recognized to Digital Innovation Hubs, *see* European Commission, *Digital Innovation Hubs: helping companies across the economy make the most of digital opportunities*, (Jan. 12, 2021). Available at: https://ec.europe.eu/digital-single-market/en/news/digital-innovation-hubs-helping-companies-across-economy-make-most-digital-opportunities. According to the European Commission, making the European Union a lighthouse center of research requires also upskilling the workforce, offering world-leading masters programs, and attracting the best professors and scientists, *see* White Paper, *supra* note 35, at p. 7.

[76] The Unfair Commercial Practices Directive 2005/29, 2005 O.J. (L 149/22) (EC); and the Consumer Rights Directive 2011/83, 2011 O.J. (L 304/64) (EC).

[77] The EU legislative framework protecting against discrimination encompasses the Race Equality Directive 2000/43/EC, the Directive on equal treatment in employment and occupation 2000/78/EC, the Directives on equal treatment between men and women in relation to employment and access to goods and services 2004/113/EC and 2006/54/EC. In addition, as from 2025, the Directive (EU) 2019/882 on the accessibility requirements for products and services will apply. It is noteworthy that the Commission's Advisory Committee on Equal Opportunities for Women and Men is expected to publish by the end of the year an Opinion on Artificial Intelligence analyzing, among other things, the impact of AI on gender equality. In fact, AI risks intensifying gender inequalities, as stated in the Communication of the European Commission, *A Union of Equality: Gender Equality Strategy 2020-2025*, COM (2020) 152 final (Mar. 5, 2020). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0152&from=EN.

framework should effectively balance protection and innovation, while not being excessively prescriptive and burdensome for businesses.

III. The Proposed AI Liability Framework

To fil the gap in the EU existing liability regime that AI specificity generated, the EU institutions are currently working to establish a modernised, specific regulatory framework for liability. Accordingly, in April 2021, the Commission has presented its AI package.[78] Beside the Communication on fostering a European approach to AI[79] and the Review of the Coordinated Plan on AI (included in the Annexes to the Communication),[80] the proposed AI Act, coupled with a new Machinery Regulation, adapts the rules to address the emerging risks and challenges posed by AI to machinery products and seeks to ensure that they can be operated safely.[81] In addition, the initial AI package has also been recently complemented with the adoption of two new legislative instruments, e.g. the proposal for a revised Product Liability Directive[82] and a proposal for an AI liability directive.[83] The combination of these proposals aims at establishing a comprehensive AI liability framework.

In this section we analyze the proposed rules for the achievement of an AI liability framework which encompasses both preventive and compensatory measures. With the AI Act and the Machinery rules on the side of ex ante regulation, and the combination of the revised PLD and the new AILD on the side of ex post regulation, we end up with a bipartite model. In the following we focus in particular on the liability directives as they introduce sets of rules that, by encompassing different types liability and for different harms, aim to be truly complementary by building an overall effective civil liability system that pushes for increased trust in AI, enhances legal certainty to encourage investments in innovation, and guaranties fair compensation for harm if it were to occur in spite of the preventive AI Act requirements.

[78] AI Package, *supra* note 16.
[79] Communication on fostering a European approach to AI, *supra* note 16.
[80] 2021 Review of the Coordinated Plan on AI, *supra* note 16.
[81] *See infra*, section III.A.
[82] Revised PLD, *supra* note 17.
[83] AILD, *supra* note 18.

A. The AI Act and the Machinery Regulation

Harmonizing the rules that govern AI systems is deemed necessary by the EU in order to ensure the safety of such systems, foster trust and legal certainty, and establish safeguards against non-compliance. To this end, the EU AI package encompasses two Regulation proposals: the AI Act[84] and the Machinery Regulation[85], both setting out obligations for economic operators of certain AI systems or – in the case of the Machinery Regulation – of machinery products incorporating AI systems.

The proposed AI Act, and the requirements it sets out, seeks to guarantee the safety of AI systems – along with other safety rules[86] – reduces the associated risks, and compels organizations to develop and use such systems in a manner that prevents the occurrence of related damage.

To do this, it differentiates between AI systems that are prohibited because they create unacceptable risk (i.e. systems meant for real-time remote biometric identification, social scoring, or manipulative AI systems, as per Article 5)[87] from those that are high-risk ("HRAIS") (such as AI systems used for credit scoring, or recruitment and for determining access to education),[88] and those that present low or minimal risks (such as chatbots). In particular, the risks taken into consideration relate to the impact that the use of an AI system can have on fundamental rights and Union values. The proposal aims at minimizing risks by introducing a system of obligations depending on the level of danger that a system presents.[89]

Therefore, the AI Act imposes substantive and procedural requirements for HRAIS aimed at enhancing, inter alia, accountability and transparency (through documentation and logging obligations), accuracy, fairness and safety (through human oversight, conformity-assessment and data quality obligations), and robustness (through effective cybersecurity and risk management). It thus allocates duties between the

[84] AI Act, *supra* note 16.
[85] Machinery Regulation, *supra* note 57.
[86] Sectoral safety rules include, among others, Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019 on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users [...], OJ L 325/1 (Vehicle General Safety Regulation); Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices [...], OJ L 117/1 (Medical Device Regulation).
[87] For more on systems presenting unacceptable risk (manipulative systems, social scoring and biometric systems), see: Veale, M., and Zuiderveen Borgesius, F. (2021). "Demystifying the Draft EU Artificial Intelligence Act – Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach", *Computer Law Review International* 22, pp. 97-112.
[88] AI Act, *supra* note 16, at Article 6 and Annexes II and III.
[89] Schuett, J. (2023). "Risk Management in the Artificial Intelligence Act", *European Journal of Risk Regulation*, pp. 1-19. Available at: https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/risk-management-in-the-artificial-intelligence-act/2E4D5707E65EFB3251A76E288BA74068

different types of operators involved in the supply chain (design, development and deployment) of HRAIS.[90] Such obligations fall *primarily* on providers (Articles 16 to 23), sometimes on product manufacturers (Article 24), and — additionally or alternatively — on users (Articles 28-29). Similarly to what is established by the product safety regime,[91] there is also the need to ensure control over dangerous products entering the EU. Consequently, importers and distributors are bound by obligations under specific circumstances (Articles 26–28),[92] with authorized representatives potentially playing a crucial role (Article 25). As to AI systems that are less risky – namely those systems meant for interaction with natural persons, emotion recognition or biometric categorization, generation and manipulation of 'deep fakes' – transparency obligations are set out for providers and users (Article 52). Moreover, all operators can be required to provide access to data and documentation to authorities enforcing the law-at-hand (Article 64).

In addition to legal certainty, another objective of legislators is the establishment of a safety-related framework that fosters confidence and is fit for "the development, use, and investment in AI systems" [93] – hence the need for a new Machinery Regulation proposed in parallel with the AI Act. As a matter of fact, the EU mechanical engineering industry continuously integrates technological advancements, ranging from advanced materials to advanced manufacturing – such as manufacturing 3D-printers, the use of robotics and autonomous machinery.[94] This development makes crucial that such markets address the new risks and ethical implications stemming from the use of AI and EDTs in general.[95] Therefore, the aim of the new Machinery Regulation is to establish effective market surveillance, legal certainty, trust in (and adoption rate of) new technologies, innovation, and competitiveness of the machinery sector – both in the European single market and globally.[96]

---

[90] This top-down framework tends to be more fit for tangible products, rather than for upstream AI services "which can be re-used downstream in a wide variety of unforeseeable contexts" (Edwards, L. (2022) "Expert explainer: The EU AI Act proposal", *Ada Lovelace Institute*. Available at: https://www.adalovelaceinstitute.org/resource/eu-ai-act-explainer/).

[91] *See infra*, Section III.B.

[92] Distributors, importers, users, as well as any other third-party, have additional obligations when they modify the substance or purpose of HRAIS, or when it is placed under their name or trademark. (Article 28 AI Act). Questions arise as to what is considered a "substantial modification" under Article 3(23) of the AI Act.

[93] European Commission Press Release (2022), *Commission welcomes political agreement on new rules to ensure the safety of machinery and robots*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7741.

[94] Lusardi, G. and Angilletta, A. (2022) "The interplay between the new Machinery Regulation and Artificial Intelligence, IoT, cybersecurity and the human-machine relationship", *Technology's Legal Edge, DLA Piper,* 29 April [online]. Available at: https://www.technologyslegaledge.com/2022/04/the-interplay-between-the-new-machinery-regulation-and-artificial-intelligence-iot-cybersecurity-and-the-human-machine-relationship/.

[95] Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee, on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64 final. Available at: https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX%3A52020DC0064.

[96] European Commission Press Release, *supra* note 99.

To ensure the safety of the next generation of machinery, the new Machinery Regulation sets out new obligations distributed between all economic operators of the supply and distribution chain — in a way that is clear, proportionate and corresponding to the role of each operator. These economic operators include manufacturers (Article 10), importers (Article 12) and distributors (Article 13); and authorized representatives can also be granted a crucial role (Article 11). To guarantee the presence of solely safe products on the market, all the above market players will have to fulfil essential health and safety requirements for machinery (whether consumer products or industrial), including AI systems encompassed in machinery. They are thus bound by obligations relating to transparency and accountability, ranging from accurate and updated documentation (such as instructions for use), record-keeping on their operations (for traceability), and participation to authorities' market surveillance tasks (by providing information) (distributors and importers).

Moreover, autonomous and "self-learning" robots are accompanied, to some extent, by higher risks than mere AI software.[97] Their mobility can not only increase the possibility of physical harm, but also harm for mental health as well as psychological tension that comes from confronting such entities.[98] The new Machinery Regulation complements the AI act in governing advanced robotics by addressing the case in which AI systems are incorporated into machines.[99] As a result, robots that are guided by AI systems and defined as "machines" need to fulfil both the essential requirements of the AI act and the new Machinery Regulation. In particular, the Machinery Regulation applies to ensure the safe integration of an AI system into a machine, to avoid compromising the safety of the machinery product as a whole:[100] it does so by requiring a conformity assessment[101] and introducing a CE marking – which indicates conformity with relevant Union requirements.[102] This internal regime of AI conformity to the relevant requirements does not in any way prevent the implementation of external controls by competent authorities nor the application of liability regimes in the event that obligations are violated.[103] Moreover, the risk posed by a certain

[97] Ragonnaud, G. (2023) "Revision of the Machinery Directive (REFIT) Q2 2021", *European Parliament Legislative Train 01.2023: A Europe Fit for the Digital Age*. Available at: https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-revision-of-the-machinery-directive.
[98] Lusardi, G., and Angilletta, A., *supra* note 94.
[99] European Commission Press Release, *supra* note 99.
[100] Machinery Regulation, *supra* note 57, at Recital 19.
[101] Id., at Article 3(28).
[102] Id., at Article 3(25).
[103] Castets-Renard, C. and Besse, P. (2022) "Ex ante Accountability of the AI Act: Between Certification and Standardization, in Pursuit of Fundamental Rights in the Country of Compliance", in Castets-Renard, C., and Eynard, J. (eds.) *Artificial Intelligence Law:*

machinery product has to be established based on "the combination of the probability of occurrence of harm and the severity of that harm",[104] and the machinery product's risk assessment has to account for the AI system's risk assessment as per the AI Act.[105]

In both regulations, infringements of the requirements and obligations would lead to penalties, including administrative fines (Article 71) in the case of the AI Act and criminal sanctions in the case of the Machinery Regulation (Article 48). However, such penalties are first established at national level by Member States, and implemented such that they are effective, proportionate, and dissuasive (as well as accounting for small-scale providers and start-up). Introducing penalties against parties for a lack of compliance for the obligations proposed by the AI Act and the Machinery Regulation can be viewed as a way to hold them responsible for their handling of such systems.[106] Even though a lack of compliance with the proposed risk-management obligations is a way to make them indirectly liable for the harm that the AI deployment generates, the question still arises as to who should compensate the AI-caused harm that materializes once the risk is not managed. The AI Act and the Machinery Regulation do not provide status recognition, procedural rights (e.g. to seek redress), nor complaint mechanisms (i.e. right to an effective remedy and a fair trial) for victims of harm caused by AI systems.[107] Therefore, holding economic operators liable imply a step further from the proposed new rules on AI governance and machinery safety, a gap in private enforcement that is somewhat filled by the so called "liability proposals".[108]

B. The proposals for a revised Product Liability Directive and …

In contrast to the AI and Machinery regulations' preventive scope, the proposed revised PLD and new AILD have a compensatory scope, aimed at ensuring that victims can effectively obtain compensation if they have suffered AI-related damages that occur despite the preventive measures required under the new

Between Sectoral Rules and Comprehensive Regime. Comparative Law Perspectives. Bruylant, Forthcoming. Available at: https://ssrn.com/abstract=4203925.

[104] Machinery Regulation, *supra* note 57, at Article 5(3).

[105] Machinery Regulation, *supra* note 57, at Recital 29.

[106] In particular, the AI Act often refers to the concept of responsibility in various provisions – namely in recitals 53 and 58, and articles 24, 26,27, and 48.

[107] Castets-Renard, C. and Besse, P., *supra* note 103, at p. 23.

[108] *See infra*, sections III.B and III.C.

AI Act, the novel Machinery Regulation, and existing sectoral safety rules (for instance on road vehicles[109] or medical devices).[110] By providing distinct yet overlapping means for redress in the occurrence of AI-related harm, these rules aim at promoting legal certainty for businesses and public trust in AI technology (as well as other EDTs).[111] While seeking to establish liability, they also facilitate the quest for information about AI systems and indirectly push for transparency in line with the *ex ante* rules above analyzed.

In particular, as to the revised PLD,[112] the aim is to modernize existing rules on strict liability on producers for defective products and ensure legal certainty, redress and fair compensation by focusing only on certain types of harm (mainly suffered by individuals), i.e. harm caused by defective products including digital and refurbished ones, so to align the product liability regime to the digital environment.[113] As a matter of fact, the original PLD, adopted in 1985, presents various gaps with respect to the increasingly digital, circular, and global economy — specifically it does not address EDTs, AI systems, and the use of any type of software embedded in products. These limitations have given rise to the need for a set of new rules, to be implemented into national law, with a much broader scope of application than its predecessor.

Hence, the proposal for a revised PLD introduces several changes to the previous one in relation to both its subjective and objective scope. As to the former, its application extends beyond producers, to address all the economic operators listed in Article 7 as parties that can be held liable for defective products adopting a "layered approach".[114] This means that compensation is sought at the next layer if parties from the previous layers cannot be held liable because they are not established in the EU or cannot be identified.[115] Liable parties will then include, beside the manufacturer of the defective product or component (or the provider of a related service), its authorised representative, the importer of the product, the so-called 'fulfilment service providers', or — under certain conditions — each distributor of the product, also any 'natural or legal person that modifies a product that has already been placed on the market or put into

[109] Vehicle General Safety Regulation, *supra* note 92.
[110] Medical Device Regulation, *supra* note 92.
[111] Revised PLD, *supra* note 17, at Section 1.1. of the Explanatory memorandum; AILD, *supra* note 18, at Section 1.1. of the Explanatory memorandum.
[112] Revised PLD, *supra* note 17.
[113] European Commission Press Release (2022), *New liability rules on products and AI to protect consumers and foster innovation*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_5807.
[114] De Bruyne, J., Dheu, O., and Ducuing, C. (2022) "The European Commission's Approach to Extra-Contractual Liability and Ai – An Evaluation of the AI Liability Directive and the Revised Product Liability Directive", *CiTiP Working Paper Series 2022*. Available at: https://ssrn.com/abstract=4295676.
[115] Revised PLD, *supra* note 17, at Article 7.

service' to be held liable (as a manufacturer), but only for 'substantial' modifications "undertaken outside the original manufacturer's control".[116]

As to the objective scope, the proposed notion of product encompasses "any movable even if integrated into another movable or into an immovable", so to include a direct reference to intangible items such as digital manufacturing files and software.[117] The updated PLD would therefore also cover AI systems and AI-enabled goods, thereby victims of damage caused by their deployment will be compensated without having to demonstrate fault, but rather on the basis of a product defect.

Also the notion of defect and damage are updated. Under the revised PLD, damage refers to material losses resulting from three different causes. In addition to the traditional damages for "death or personal injury" (including to psychological health), and "harm to or destruction of property", also the "loss or corruption of data that is not used exclusively for professional purposes" may now require compensation.[118] The notion of defect, instead, somewhat still resembles the one adopted under the old PLD, however, the non-exclusive list of circumstances to consider in this assessment is expanded. Article 6(1)(c), for instance, allows the assessment of a defect to account for "the effect on the product of any ability to continue to learn after deployment". As such, the product liability regime seeks to be "future-proofed" against additional improvements in self-learning AI systems that would keep developing post-deployment.

Against this background, the proposal for a revised PLD encompasses two provisions that are key with respect to the harm caused by AI products. In the first place, Article 8 empowers national courts to order a defendant to disclose relevant evidence, within the limits of what is "necessary and proportionate to support a claim" (the meaning of which is left to the court's discretion). This can be ordered at the request of a claimant, provided that they present "facts and evidence sufficient to support the plausibility of the claim for compensation." Non-compliance with the disclosure order would lead to a presumption of the defect. In this way the opacity of AI systems could be overcome by such an order.

---

[116] Revised PLD, *supra* note 17, at Article 7(4).
[117] Revised PLD, *supra* note 17, at Article 4(1).
[118] Revised PLD, *supra* note 17, at Article 4(6). See *e.g.* Twigg-Flesner, C., *et al.* (2021) "Guiding Principles for Updating the Product Liability Directive for the Digital Age", *ELI innovation paper series*, p. 8, where it is clearly stated that "[r]evisions of the notion of 'damage' could be considered to include damage to digital elements and data."

In the second place, Article 9 governs the burden of proof and the conditions for claiming compensation – i.e. proving the product defectiveness, the damage suffered, and the causal link between the two (Article 9(1)). In this context the revised directive establishes two rebuttable legal presumptions. First, the presumption of the product's defectiveness (in Article 9(2)) is triggered under certain conditions that make up three possible scenarios: (*i*) when the defendant fails to disclose relevant evidence (and to comply with the disclosure order above mentioned) for instance on system information; (*ii*) when the claimant proves that the product violates (EU or national) mandatory safety requirements specifically meant to avoid this damage; or (*iii*) when the claimant shows that the damage arose from the product's "obvious malfunction" under normal use or "ordinary circumstances". Second, the presumption of causality, i.e. of causal link between the product's defectiveness and the damage (in Article 9(3)), activates when "it has been established that the product is defective and the damage caused is of a kind typically consistent with the defect in question". This can be particularly burdensome in the case of opaque, black-box AI systems, where it is difficult, if not impossible, to understand their functioning and thus their probable contribution to the damage.

In addition, Article 9(4)) introduces a more general presumption that is also crucial for AI products as it allows both of the above to work simultaneously. When a national court finds that "the claimant faces excessive difficulties, due to the technical or scientific complexity",[119] to establish the product's defectiveness and/or the causal link with the damage', either or both can be assumed under certain conditions. This aims to ease claimants' burden of proof in complex situations where products violate safety requirements. However, for the above presumption to apply, the claimant must still prove, based on "sufficiently relevant evidence", the product's contribution to the damage, as well as the likelihood of the product's defectiveness and/or the likelihood of its defectiveness causing the damage. The defendant is empowered to rebut any of the Article 9 presumptions. In the latter case, it can be done by contesting the existence of excessive difficulties or the mentioned likelihood.

Among the defences that are available in the revised PLD for economic operators to escape liability, they cannot invoke, according to Article 10, the software or its upgrade or update, or the lack of software updates

[119] Cases of technical and scientific complexity are further clarified in Section 1.1 of the explanatory memorandum as "those involving pharmaceuticals, smart products or AI-enabled products (…)" (Revised PLD, *supra* note 17).

or upgrades necessary to maintain safety if they are within the manufacturer's control. In fact, the revised directive recognises that software and AI systems are updatable after having been placed on the market. Such updates, upgrades, and related services can make a product defective even if it was not defective when put into circulation, thus causing harm for which compensation can be claimed. Manufacturers can even be held liable when failing to provide "software updates or upgrades necessary to maintain safety", like to overcome cybersecurity vulnerabilities. In assessing defectiveness and liability, it is relevant to see whether manufacturer have control over the product, independently of whether it has already been placed on the market (as long as it was not before the directive's transposition). Moreover, the rights conferred pursuant to the revised PLD extinguish 10 years after products have been put into circulation: this long-stop is extended in case of any "substantial modifications" to the product, likely including software updates.

In summary, the revised PLD enlarges the strict liability regime to software and, therefore, AI-products, and aligns its provisions to their functioning. Given this extended – and to a certain extent even *new* regime – it is likely that insurers will increase scrutiny over economic operators falling under the revised PLD's scope, specifically over their rules governing AI, software, cybersecurity.[120]

C. … a new AI Liability Directive

By providing rules to compensate harm caused during the use of AI systems, the proposed AILD aims to first foster innovation in the AI sector, by reducing uncertainty for players that operate in several jurisdictions, and increasing guarantees (through instruments such as the right to rebut a liability claim based on a presumption of causality). Second, to increase consumers' trust when interacting with AI, by enhancing their protection, up to the "same standards of protection when harmed by AI systems as they would be if harmed under any other circumstances."[121]

[120] Kidman D., et al. (2022). "The Proposed New Product Liability Directive: Reviving the consumer friendly approach of the existing directive, for the 21st century", *Simmons+Simmons*. Available at: https://www.simmons-simmons.com/en/publications/cl940urlm5m0y0b797165na9z/the-proposed-new-product-liability-directive.

[121] "Questions & Answers: AI Liability Directive" (2022), *European Commission – Questions and Answers*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/QANDA_22_5793.

In order to enable a claimant to substantiate their claims, these rules are supposed to harmonize access to information and disclosure of evidence on HRAIS, and to significantly alleviate the burden of proof for victims to access compensation. Indeed, this instrument intends easing access to redress by removing additional evidentiary hurdles and facilitating the legal process for claimants to prove that the defendant's fault – arising from an act or omission resulting from the use of AI systems – led to damage.

In concrete, the AILD proposes a mechanism for national claims of fault-based liability for damages caused by AI systems (intentionally or negligently, as clarified in Article 1 defining the AILD's scope). To do this it provides the means to prove someone's fault in the occurrence of AI-related harm, and to offer potential compensation for victims of such damage (for all types of damage and victim). This complements the revised PLD as it addresses damages that are not caused by defect,[122]and range from breaches of privacy to damages caused by safety issues to discrimination in recruitment processes involving AI-systems.

However, unlike the revised PLD (which is based on maximum harmonization), the AILD is based on the minimum harmonization approach, which allows Member States to introduce or maintain domestic laws that ensure additional or further support to claimants when it comes to AI-caused damage, for example regarding strict liability regimes and reversals of the burden of proof. Indeed, the directive provides Member States with enough leeway on the implementation of the AI liability regime into domestic law which is specific to each jurisdiction (for instance regarding the understanding of fault and damage) and emanates from long-established legal traditions.

In order to avoid overstepping on national civil liability law, the proposed directive requires Member States to adopt only two main changes to its law — namely two presumptions set out in Articles 3 and 4.

The first of these presumptions is linked to the preventive measures adopted in the AI Act. As known, the AI Act introduces specific obligations for high-risk AI systems. Complementary to those obligations, Article 3 of the AILD introduces an evidence disclosure mechanism and a rebuttable presumption of non-compliance with the duty of care. Specifically, to support liability claims of persons injured by AI systems, Article 3 allows them to seek information through national courts. The latter are empowered to order that

[122] The AILD is also aligned to the AI Act: for instance, the AILD definitions (Article 2(1-4)) mirror the AI Act's definitions as for "AI-system" (Article 3(1) AI Act), "High-risk AI system" (Article 6 AI Act), "provider" (Article 3(2) AI Act) and "user" (Article 3(4) AI Act). This confirms the proposition that they both contribute to build the EU AI liability regime.

a party bound by AI Act obligations – e.g. HRAI providers (or parties bound by their obligations), HRAI users, persons in possession of relevant information, etc. – should disclose certain relevant evidence at their disposal regarding a specific HRAI-system suspected of having resulted in the damage. Disclosure should be limited to the information that is necessary and proportionate to support the liability claim. The AILD also provides the possibility for a claimant to request both the disclosure and the preservation of evidence (Article 3(3)). Such disclosure could be ordered by courts both during and before the initiation of proceedings on the merits, although with conditions that vary depending on the timing.

On the other hand, a claimant who wishes to obtain a disclosure order must first engage in every proportionate effort to gather the relevant evidence from the defendant. Potential claimants can request disclosure orders before initiation of proceedings on merits only if they had first requested such disclosure to the provider (or person subject to its obligations, or user) but it was refused, and if they provide enough facts and evidence to support a plausible claim (in order to avoid 'fishing expeditions'). Prior disclosure would assist in the identification of relevant evidence to support the claim and of potentially-liable actors (thus also eliminating those incorrectly identified, and limiting unnecessary litigation).[123]

In any case disclosure is subject to safeguards, such as limiting access to minimum and proportionate information, in order to "prevent blanket requests". When assessing the proportionality of disclosure or preservation orders, the proposed AILD requires national courts to take into account the legitimate interests of all parties (among which providers and users) and confidential information (Article 3(4)). When the order concerns an (alleged) confidential information or trade secret,[124] and when a party or ex-officio submits a duly motivated request, national courts are empowered to conduct the balancing exercise between disclosure/preservation and protection of secrecy, and to adopt specific measures necessary to ensure confidentiality. These measures can include, as provided by Directive 2016/943 on trade secret, redacting sensitive portions of rulings and restricting the number of individuals granted access to some evidence.[125]

[123] Counseson, G., Hendrix, G.J., Bellon, J. (2022). "EU – Taking responsibility for artificial intelligence: New tort liability proposals", *DigiLinks, Linklaters*. 3 October. Available at: https://www.linklaters.com/en/insights/blogs/digilinks/2022/october/eu---taking-responsibility-for-artificial-intelligence_new-tort-liability-proposals.

[124] These are specifically referenced in the AI Liability Directive, under the Trade Secret Directive (EU Directive 2016/943) and national transposing legislation. *See* AILD, *supra* note 18, at Article 3(4).

[125] This mechanism closely resembles Article 13 of the AI Act. However, while the latter sets out a transparency requirement focused on providing information to the user of the AI-system, Article 3 AILD sets out the disclosure of evidence to any victim

In a claim for damages, if the defendant fails to comply with a court order to disclose or preserve evidence at its disposal (for instance because it never arranged to document or preserve it), an easing of the burden of proof applies. Indeed, under the new regime, national courts can invoke a rebuttable presumption of non-compliance with the defendant's relevant duty of care (under EU or national law) that the requested evidence was meant to prove (Article 3(5)). However, the presumption only applies if the court finds excessively difficult for the claimant to prove the causal link, and so long as three specific conditions are satisfied: (*i*) it must be proved by the claimant[126] or presumed by the court[127] that the defendant (or someone whose behaviour is the defendant's responsibility) has committed a relevant fault;[128] (*ii*) it should be considered reasonably likely, in light of the circumstances, that the defendant's fault has affected the output produced by the AI system or its failure to do so; (*iii*) it is demonstrated by the claimant that the output produced by the AI system or its failure to do so gave rise to damage.

The second presumption is encompassed in Article 4 and alleviates the claimant's burden of proof by introducing a rebuttable presumption of causality that infers the causal link between the defendant's fault and the AI system's produced output (or failure to do so) that gave rise to the damage. Although its scope is limited, the presumption helps claimants in AI-related liability cases to overcome the difficulties – exacerbated by the complexity and autonomy of AI systems – faced in providing such evidence. The presumption, however, only applies if the requirement of fault (the breach of duty) is established by the court or proved by the plaintiff.

In the case of HRAIS, it varies between claims brought against producers (or a person subject to providers' obligations under the AI Act) and those brought against users. In the former case, Article 4(2) provides that the presumption applies when the provider: (*i*) fails to comply with the obligations (exhaustively enumerated in the AILD, and directly stemming from the AI Act) that relate to the quality of data (training and testing of datasets), "transparency, human oversight, system accuracy, robustness and cybersecurity"; or (*ii*) fails to take corrective actions to remedy another breach or recall a HRAIS when the fault was

of AI harm. The AILD facilitates the process of demonstrating fault by relying on the disclosed evidence, thus supporting damage claims for compensation.

[126] Such proof would have to be established by the claimant according to the applicable EU law or national rules.

[127] This would follow AILD Article 3(5): *See* AILD, *supra* note 18, Article 3(5).

[128] This fault would consist in a breach of duty of care (i.e. an obligation under national or EU legislation, such as the AI Act) that is directly intended to protect against the harm that occurred.

identified. In the latter case, Article 4(3) provides that it applies when the user either fails to comply with its related AI Act obligations 'to use and monitor an AI system in accordance with accompanying instructions of use or, where appropriate, suspend or interrupt its use' (Article 29 AI Act); or exposes the AI system to input data under its control which is not relevant in view of the system's intended purpose (Article 29.3 AI Act).

However, the AILD also introduces limitations to the application of the rebuttable presumption of causal link that vary according to the nature of the AI used.[129] For HRAIS, the court does not apply the presumption when the defendant demonstrates that sufficient evidence and expertise is reasonably accessible for the claimant to prove the causal link – namely through the AI Act obligations related to transparency, documentation, logging and recording. This specific exception related to HRAIS (Article 4.4) can prompt defendants' compliance with such requirements. For standard non-HRAIS, instead, the court only applies the presumption where it considers it "excessively difficult for the claimant to prove the causal link" — which is determined based on certain AI systems' characteristics (autonomy, opacity, etc.).

In addition, there is a more general way to rebut the burden of proof. To further incentivize disclosure, the AILD provides (Article 4.7) that if the causality presumption is invoked, the defendant may rebut it by demonstrating that "sufficient evidence and expertise is reasonably accessible for the claimant to prove the causal link". However, it is likely difficult to provide such sufficient evidence, i.e. that the damage suffered could not have been caused by the fault (evidence of a negative fact) or that it has been caused by another factor (requiring a complete view on the facts).

In conclusion, the AILD is to welcome even though it does not shift the burden of proof on the defendant but it only alleviates it. Some scholars consider that a complete shift would be excessively burdensome,

[129] For AI systems used in the context of personal non-professional activities, the AILD (Article 4(6)) sets out yet another differentiated regime: the causality presumption applies only when the defendant "has materially interfered with the conditions of the operation of the AI system", or was both required and able to determine such conditions yet did not. Non-professional users of AI systems whose behaviours do not add risk are exempted from the presumption; this aims at balancing their interests with those of victims. *See* Lusardi, G., and Darling, C. (2022) "The AI Liability Directive: EU Improves Liability Protections for Those Impacted by AI", *Technology's Legal Edge, DLA Piper*, 6 December. Available at: https://www.technologyslegaledge.com/2022/12/the-ai-liability-directive-eu-improves-liability-protections-for-those-impacted-by-ai/#page=1.

potentially put up barriers to innovation[130] and to the adoption of AI-systems, and increase contentiousness and litigation against several potentially-liable parties.[131]

CONCLUSION

In this article, we have analyzed the "two sides of the same coin", namely the rules on AI governance and on AI liability as they both contribute to creating an AI liability framework in the EU. Compliance with the obligations introduced under the AI Act, as well as with all the obligations requiring safety measures to be in force – such as those under the proposal for a new machinery regulation – is essential to enhance trust in AI within the digital single market. Compliance efforts can include system inventories, assessment of required steps, risk assessments, and appropriate governance that is proportionate to the risk generated by the use of AI in the specific context. The proposals for the AILD and the revised PLD reinforce the importance of compliance with these obligations, particularly concerning HRAIs, and provide tools for consumers to be compensated if compliance is not enough to prevent harm. This balance between enhancing consumer trust and protection and fostering innovation investment is essential to ensure that AI benefits society.

However, some uncertainty persists as to how the various provisions that constitute the liability framework should be coordinated. In the first place, there is a potential overlap among the different pieces of legislation, particularly the AILD and the PLD. In principle, the fault-based regime introduced by the former should not overlay the no-fault regime of the latter. The AILD clearly states that it does not affect "any rights which an injured person may have under national rules implementing" the existing PLD,[132] which however differs significantly from the revised version just proposed. The relationship between the AILD and the revised PLD thus requires further clarification, for instance on the concept of burden of proof, as there could be scenarios where the injured party is relying on both the PLD and the AILD to seek

[130] De Bruin, R. (2016) "Autonomous Intelligent Cars on the European Intersection of Liability and Privacy", *European Journal of Risk Regulation* 7(3), p. 495.
[131] Schütte, B., *et al.* (2021). "Damages Liability for Harm Caused by Artificial Intelligence – EU Law in Flux", *Helsinki Legal Studies Research Paper* 69, p. 26.
[132] AILD, *supra* note 18 at Article 1.3.(b).

compensation for harm caused by an AI system. In such cases, it is unclear how the burden of proof would be allocated between the two liability regimes. The question arises as to whether the presumption of defect established under the PLD would still apply, or the injured party would need to prove fault – even though alleviated – under the AILD.

In the second place, doubts also arise as to the delimitation of the two directive proposals' respective scopes.[133] For instance, consider the case of companies which use standard or complex algorithms that do not fall under the strict definition of AI,[134] such as the algorithms often used in the financial sector. Indeed, complex algorithms (or software) may not use machine learning or other AI techniques, yet they can still generate significant risks and have unintended consequences that could harm consumers. While such software would fall under the scope of the revised PLD, it may not necessarily fall under the scope of the AILD, thereby creating confusion for potential plaintiffs as to which compensation regime to rely on. In addition, the revised PLD does encompass all software but sets aside specific provisions (on the reversal of the burden of proof) solely for software that exceeds a certain level of complexity. However, these complex algorithms can create risks of unforeseeability and opacity, independently of whether they qualify as AI as defined in the legislation[135] – a consideration with potential importance for the AI Act and the AILD, which adopt a risk-based perspective.[136]

Further uncertainty accompanies the choice of directives as legal instruments, for the unclarity and lack of precision surrounding certain notions that could be applied differently according to national law and to national courts' interpretation and discretion. This could result in legal uncertainty, especially when national traditions adopt diverging approaches, as it is often the case in tort law.[137] For example, in the AILD, notions such as "fault" and "user" are left to interpretation. There is also margin for subjectivity regarding certain requirements for the application of presumptions.[138] It is indeed unclear what constitutes "all

[133] Hacker, P. (2022) "The European AI Liability Directives – Critique of a Half-Hearted Approach and Lessons for the Future". Available at: https://ssrn.com/abstract=4279796.
[134] Bruner, G. (2020) "No, Artificial Intelligence doesn't exist (yet)", *Towards Data Science*, 28 December. Available at: https://towardsdatascience.com/no-artificial-intelligence-doesnt-exist-yet-3318d83fdfe8.
[135] Lipton, Z.C. (2018) "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery", *ACM Queue* 16(3).
[136] Hacker, P., *supra* note 133.
[137] De Bruyne, J., Dheu, O., and Ducuing, C., *supra* note 114.
[138] Madiega, T. (2023) "EU Legislation in Progress Briefing: Artificial intelligence liability directive", *European Parliamentary Think Tank*. Available at: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)739342.

proportionate attempts" to gather relevant evidence, which are required before disclosure of evidence can apply.[139] Eventually, as to the causality presumption,[140] there are notions that are subject to case-by-case analysis such as the requirement of "reasonable likelihood" of fault influencing the output, or the "duty of care" in relation to the concept of fault, and the exception when "sufficient evidence and expertise [are] reasonably accessible".

Similarly, there is unclarity also regarding some terms and provisions of the revised PLD, including the definition of "digital services" encompassed by "related services" (under Article 4.4), and the definition of "software update or upgrade" (under Article 4.5). Moreover, the question of whether "modifications" include self-learning capabilities of software depends on how the concept of "manufacturer's control" will be subject to possibly divergent interpretations.[141] Also the notion of defect, as defined under Article 6, remains somewhat unclear, due to undefined terminology such as "reasonably foreseeable" (when referring to product (mis)use). Finally, understanding what constitutes a "substantial modification", as per Article 7(4), also presents its fair share of challenges.

Beside the significant leeway for Member States' courts to apply their national liability regime to cases of AI-caused damage, national courts will also need to understand AI systems, their functioning and their risks.[142] This is relevant, for example, when setting a standard of conduct to apply the general standard of care to cases involving AI. Assessing whether an actor's behavior compares to that of a normal and prudent person is a task that varies according to the technological state of the art.[143] For the interpretation of product liability concepts to be adequate, national courts should also be aware of sectoral safety standards, AI characteristics (such as autonomy), as well as scientific and technical frameworks of reference.[144]

More generally, the AI liability EU framework raises further criticism in relation to the risk-based approach, the horizontal approach, and the approach to alleviate the burden of proof. First, the application of the AI

[139] AILD, *supra* note 18 at Article 3.
[140] AILD, *supra* note 18 at Article 4.
[141] De Bruyne, J., Dheu, O., and Ducuing, C., *supra* note 114.
[142] *Id.*
[143] De Bruyne, J., Van Gool, E., and Boes, A. (2022) 'Wat bracht 2022 en wat brengt de toekomst op het vlak van artificiële intelligentie en buitencontractuele aansprakelijkheid?', in T. Vansweevelt and B. Weyts (eds.), *Verslagboek van het IVde Interuniversitair Congres Aansprakelijkheids-en Verzekeringsrecht,* Cambridge: Intersentia.
[144] See *e.g.* Baker, J.E., Hobart L.N., and Mittelsteadt, M.G. (2021) 'AI for Judges. A Framework', *CSET Policy Brief*, p. 32-46, recommending initiatives aimed at fulfilling judges' competences, in addition to enhanced clarity in the legislation.

Act's risk-based approach to the AILD appears highly controversial. The AILD's material scope relies on the AI Act's definition of AI systems – an alignment deemed crucial. However, the AILD is also aligned to the AI Act's risk classification of AI systems,[145] indeed some core AILD principles (i.e. disclosure of evidence and part of the burden of proof alleviation) only apply to HRAIS. Although this alignment favors coherence between intertwined pieces of legislation, applying the HRAIS categorization to AI liability regimes runs the risk of over-inclusiveness (for instance in the case of general-purpose AI systems) and under-inclusiveness (of high-risk cases).[146] The question arises as to where this under-inclusiveness arises from. The AI Act's risk-based approach focuses on the impact on society as a whole. One may argue, however, that effective liability regimes must also consider the materialization of individually pronounced risks, even when their unequal distribution among members of society places them out of the AI Act's HRAIS classification.[147] When the system's probability of damage (and thus its aggregate risk-level or expected damage) is too low to qualify it as HRAIS under the AI Act, but its risk variance between individual victims is significant, we end up with diverging classifications of social risk versus individual risk. In this case, equating risk categories in different instruments (namely, in the AI Act and the AILD) would result in the absence of effective remedies for victims of individually pronounced risks. This is the case of autonomous vehicles, which fall under the HRAIS category[148] but not under the core AI Act obligations,[149] nor – as a result – under the related AILD provisions.[150]

Second, the adoption of a horizontal liability regime applying to a variety of sectors does not consider the fact that AI works differently according to the sector it is used in. Indeed, it runs the risk of disregarding the intrinsic differences between distinct AI applications and the issues they raise.[151] Several voices have pointed out that that "no one-size-fits-all solution can (or should) be offered" regarding liability for AI-caused damage, as it fails to recognize and overcome the challenges raised by the heterogeneity of AI uses

[145] AILD, *supra* note 18 at Articles 1, 2(1) and 2(2).
[146] Hacker, P., *supra* note 133.
[147] Hacker, P., *supra* note 133.
[148] AI Act, *supra* note 16, at Article 6.
[149] AI Act, *supra* note 16, at Articles 2(2) and 84, and Annex II Section B. HRAIS that fall under the scope of the acts mentioned in Article 2(2), such as Automated Vehicles, are only subject to Article 84 and not to the AI Act's core obligations.
[150] AILD, *supra* note 18 at Articles 3 and 4.
[151] Whittam, S. (2022) "Mind the compensation gap: towards a new European regime addressing civil liability in the age of AI" *International Journal of Law and Information Technology*, 30, pp.249–265. Available at: https://doi.org/10.1093/ijlit/eaac013.

– these voices encourage instead the implementation of a range of options, with the choice within that range to be determined by various factors".[152]

Although it might require "greater effort" from policymakers,[153] some scholars deem "more appropriate" an ad-hoc, sector-specific EU liability regime, tailored to the peculiarities of different fields (parallel to the revised PLD acting as a general rule).[154] Since liability in different sectors (e.g. transport and healthcare) needs to be regulated by different frameworks when AI is out of the picture, it might be the case also when AI is used in both sectors. Elements that would be tailored to a specific AI application could include strictly liable parties, as well as remedies and obligations of termination, non-repetition, redress, and compliance.[155]

Third, the approach toward an alleviated claimants' burden of proof is also somewhat criticized. While the ex-ante part of the AI liability framework is very articulated, the ex-post part would not be effective enough as it still puts some heavy burden of proof (and some vulnerability) on claimants. Indeed, for the directives' presumptions to apply, claimants still have to prove numerous elements such as defect (PLD) or fault (AILD) on one side, and damage on the other side, as well as the nexus between these two sides. For disclosure of evidence to apply,[156] claimants (PLD) or potential claimants (AILD) must first "present facts and evidence sufficient to support the plausibility of a claim".[157] Furthermore, for the AILD's causality presumption to apply in the case of HRAIS covered by the AI Act, the claimant must prove non-compliance with the AI Act requirements (Article 4 AILD).[158]

The coordination issues and general criticism above-illustrated reverberate on the effectiveness of the proposed AI liability European framework at various level, in particular at the level of organizations and consumers. Organizations that strongly invest in AI should start accounting for this new liability framework – despite the lack of legal certainty – and taking steps towards future compliance with its requirements,

[152] Expert Group on Liability and New Technologies, *supra* note 70, p. 36. Bertolini, A. (2020) "Study requested by the JURI committee: Artificial Intelligence and Civil Liability"', *European Parliament Think Tank*. Available at: https://www.europarl.europa.eu/thinktank/en/document/ IPOL_STU(2020)621926.
[153] De Bruyne, J., Dheu, O., and Ducuing, C. (2022), *supra* note 114.
[154] Bertolini, A., *supra* note 152.
[155] Ad hoc Committee on Artificial Intelligence, CAHAI (2020), "Feasibility Study", CAHAI(2020)23, p. 38. Available at: https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da.
[156] AILD, *supra* note 18 at Article 3, and Revised PLD, *supra* note 17, at Article 8.
[157] De Bruyne, J., Dheu, O., and Ducuing, C., *supra* note 114.
[158] Nawaz, S.A. (2022), 'The Proposed EU AI Liability Rules: Ease or Burden?', [Blog] European Law Blog, 7 November. Available at: https://europeanlawblog.eu/2022/11/07/the-proposed-eu-ai-liability-rules-ease-or-burden/.

especially at the development or procurement stage of AI systems. Otherwise, when these systems will get deployed after the implementation deadline, the new liability regime will expose them to significant risks, ranging from reputational damage to AI liability claims. To better prepare and defend themselves from such claims against their AI-systems, organizations may consider adopting some key steps aimed at detecting potential issues and mitigating risk of litigation.[159] This could include providing training to anyone involved in the design, development, implementation and operation of AI-systems covered by the legislation. Moreover, under the AILD, providers and users of HRAI may face the causality presumption, if they do not ensure appropriate security, monitoring, or interruption of use of the AI system when required under the AI Act. They can facilitate the rebuttal of such allegations that harm was caused by AI – and prove that it was caused by another factor – by enhancing their audit capabilities (i.e. through robust documentation and activity logs of, respectively, model testing and performance) and by conducting AI incident response planning and testing (through an AI tabletop exercise).

At the level of consumers, the proposed liability directives aim to facilitate claims of AI-caused damage, yet it remains complicated to prove fault when it comes to complex systems. Due to the black box phenomenon, the autonomy and complexity of AI systems, it is often difficult to understand the reason (and input) behind a certain output. It is not easy to prove that the datasets used in developing a system or that the accuracy level of that system are inadequate. Logged data, for instance, can be particularly hard and costly to interpret.[160] In such cases, information about the system, as provided for by the framework, does not always do much good. If the PLD overcomes these challenges of proving fault because it provides a no-fault mechanism, such strict liability mechanism is limited to material harm[161] (in line with what many developers push for)[162]. For instance, credit scoring cannot be challenged by relying on defectiveness and thus on the PLD, but only by proving fault under the AILD.[163] Further effort might therefore be required

[159] An example of measures to take is in "The EU AI Liability Directive Will Change Artificial Intelligence Legal Risks", *Debevoise In Depth*, 24 October 2022. Available at: https://www.debevoise.com/insights/publications/2022/10/the-eu-ai-liability-directive-will-change.
[160] Bertolini, A., *supra* note 152.
[161] Ursula Pachl, the Deputy Director of the European Consumer Organization (BEUC) has already voiced this concern while talking about the directives. See Bertuzzi, L. (2022) "The new liability rules for AI", Euractiv, 30 September. Available at: https://www.euractiv.com/section/digital/podcast/the-new-liability-rules-for-ai/.
[162] Joint Industry Letter on the PLD and AI Directive (2022), 24 August. Available at: https://ccianet.org/wp-content/uploads/2022/08/2022.08.24-Joint-Industry-Letter-on-the-PLD-and-AI-Directive.pdf.
[163] Nawaz, S.A., *supra* note 158.

for the effective facilitation of redress mechanisms. In this context, some commentators push for the framework to address victims' need for specific resources – both technical and financial – in order to support their claims.[164]

In addition, the decision to propose an AI liability framework that amounts, as far as the ex-post rules are concerned, to directives (AILD and PLD), may limit the overall harmonization effect that is pursued and confirm the fragmentation already existing among Member States as to liability regime. In particular, some commentators have also pinpointed that the PLD and AILD's mechanisms require to be further harmonized, for instance by empowering potential claimants with evidence disclosure also under the PLD.[165] In general, the effective harmonization that comes from the adoption of the depicted AI liability regime will need to be evaluated in light of the national implementation of the directives vis-à-vis the direct application of the ex-ante provisions.

[164] Madiega, T., *supra* note 138, referencing De Bruyne, J., Dheu, O., and Ducuing, C., *supra* note 114, calling for a clearer distribution of roles, better explanation of the underlying notions and the need for technical expertise and financial resources for victims that are required to prove their claims.
[165] Hacker, P., *supra* note 133.