12

Liability for Artificial Intelligence

The Need to Address Both Safety Risks and Fundamental Rights Risks

Christiane Wendehorst

I. INTRODUCTION

On 21 April 2021, the European Commission published its package of measures on a European approach to artificial intelligence (AI), consisting of a communication,[1] accompanied by an updated Coordinated Plan on AI[2] and a proposal for a horizontal regulation (Artificial Intelligence Act, AIA)[3] with nine annexes. This package is the first of three inter-related legal initiatives announced by the Commission with the aim of making Europe a safe and innovation friendly environment for the development of AI. This first initiative aims to establish a European legal framework for AI to address fundamental rights and safety risks specific to AI systems. The second initiative is the revision of sectoral and more horizontal safety legislation. A proposal for a new Machinery Regulation[4] with eleven annexes was already published on the same day as the AI package, addressing an important aspect of AI usually referred to as 'robotics', and a proposal for a new General Product Safety Regulation[5] followed soon after. Parliament and Council are currently preparing both files for the trilogues. Finally, the third initiative announced is the introduction of EU rules to address liability issues related to new technologies, including AI systems. The Public Consultation for this initiative has already been closed and a proposal is planned for the third quarter of 2022.[6] This third initiative will comprise measures adapting the liability framework to the challenges of new technologies, including AI, to ensure that victims

[1] European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Fostering a European Approach to Artificial Intelligence' COM (2021) 205 final.

[2] European Commission, 'Annex to the 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. New Coordinated Plan on AI 2021 Review' COM (2021) 205 final.

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

[4] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on Machinery Products' COM (2021) 202 final.

[5] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on General Product Safety, Amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council, and Repealing Council Directive 87/357/EEC and Directive 2001/95/EC of the European Parliament and of the Council' COM (2021) 346 final.

[6] European Commission, 'Civil Liability: Adapting Liability Rules to the Digital Age and Artificial Intelligence' https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12979-Civil-liability-adapting-liability-rules-to-the-digital-age-and-artificial-intelligence_en; this chapter was written in spring 2021, only certain sections have been updated.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

who suffer damage to their life, health, or property as a result of new technologies have access to the same compensation as victims of other technologies. In the Inception Impact Assessment, a revision of the Product Liability Directive (PLD),[7] and a legislative proposal with regard to the liability for certain AI systems are identified as policy options.[8]

Given that liability for AI and other emerging digital technologies had been on the agenda for some time, it may come as a surprise that liability legislation figures last on the agenda. An Expert Group on Liability and new Technologies was established in 2018. It was divided into two formations, one dealing specifically with the PLD and being largely dominated by stakeholders, the other – the so-called New Technologies Formation (EG-NTF) – having a broader mandate and consisting mainly of academics.[9] Only the NTF ever published an official written report,[10] which then served, *inter alia*, as a basis for the European Commission's report on the safety and liability implications of AI, the Internet of Things (IoT), and robotics[11] of 19 February 2020, which formed part of the 2020 AI package and accompanied the Commission White Paper on AI.[12]

A major driver of activities in the field of liability has certainly been the European Parliament. After its first resolution in 2017,[13] which included the much-quoted and much-criticised plea for electronic personhood,[14] the European Parliament passed another resolution on 20 October 2020 that includes a full-fledged 'Proposal for a Regulation of the European Parliament and of the Council on liability for the operation of AI systems'.[15] This proposal is certainly much more mature than the 2017 resolution and bears a striking resemblance to policy considerations made within parts of the European Commission.

Whether the Commission will follow the recommendations of Parliament or take a different approach remains yet to be seen. Because AI liability is a subject matter that might be addressed

[7] Council Directive 85/374/EEC of 25 July 1985 on the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Concerning Liability for Defective Products [1985] OJ L 2010/29; see European Commission, 'Commission Staff Working Document. Evaluation of Council Directive 85/374/EEC of 25 July 1985' SWD (2018) 157 final.

[8] European Commission, 'Adapting Liability Rules to the Digital Age and Artificial Intelligence' Inception Impact Assessment (Ares(2021)4266516).

[9] European Commission, 'Register of Commission Expert Groups, Expert Group on Liability and New Technologies (E03592)' (*European Commission*, 9 March 2018) https://ec.europa.eu/transparency/expert-groups-register/screen/expert-groups/consult?do=groupDetail.groupDetail&groupID=3592&Lang=NL.

[10] Directorate-General for Justice and Consumers, 'Liability for Artificial Intelligence and Other Emerging Digital Technologies' (*European Commission*, 27 November 2019) https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en/format-PDF (hereafter 'NTF Expert Group').

[11] European Commission, '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.

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

[13] European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics, P8_TA (2017)0051 (hereafter EP Resolution on Civil Law Rules on Robotics).

[14] G Wagner, 'Robot Liability' in S Lohsse, R Schulze, and D Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (2019) 44 *et seq*; BA Koch, 'Product Liability 2.0: Mere Update or New Version?' in S Lohsse, R Schulze and D Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (2019) (hereafter Koch, 'Product Liability 2.0: Mere Update or New Version?'); G Spindler, 'Roboter, Atomation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien?' (2015) CR 766, 773; H Eidenmüller, 'The Rise of Robots and the Law of Humans' (2017) ZEuP 765, 774 *et seq*; R Schaub, 'Interaktion von Mensch und Maschine' (2017) JZ, 342, 345.

[15] European Parliament Resolution of 20 October 2020 with Recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence (2020/2014(INL)) P9_TA(2020)0276 (hereafter EP Resolution on a Civil Liability Regime for AI).

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

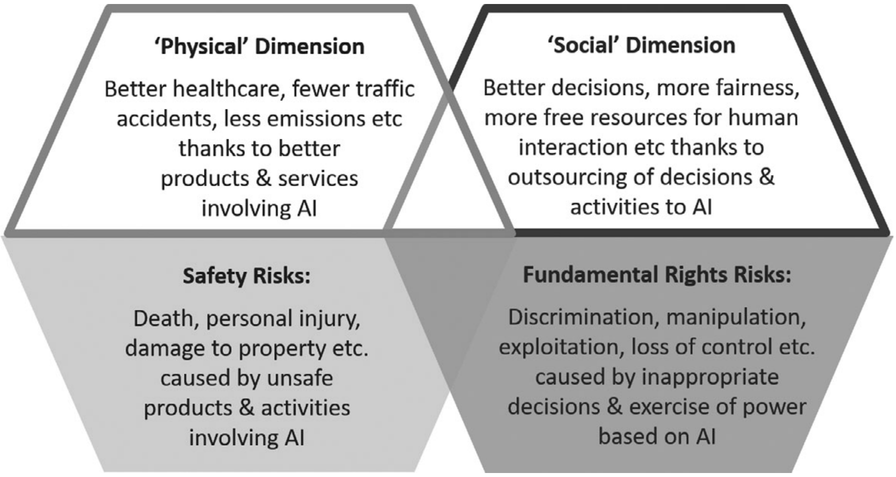
'Physical' Dimension
Better healthcare, fewer traffic accidents, less emissions etc thanks to better products & services involving AI
'Social' Dimension
Better decisions, more fairness, more free resources for human interaction etc thanks to outsourcing of decisions & activities to AI
Safety Risks:
Death, personal injury, damage to property etc. caused by unsafe products & activities involving AI
Fundamental Rights Risks:
Discrimination, manipulation, exploitation, loss of control etc. caused by inappropriate decisions & exercise of power based on AI

FIGURE 12.1 The 'physical' and the 'social' dimensions of risks associated with AI

within different regulatory and legal frameworks for which different Directorates General of the Commission and different Committees within the Parliament are responsible, the matter remains highly controversial. This paper analyses the different risks posed by AI, and why AI challenges existing liability regimes. It also explains the main solutions put forward so far and evaluates them, concluding that different solutions may be appropriate for different types of risk.

II. DIMENSIONS OF AI AND CORRESPONDING RISKS POSED

The challenges posed by AI and modern digital ecosystems in general – such as opacity ('black box-effect'), complexity, and partially 'autonomous' and unpredictable behaviour – are similar, irrespective of where and how AI is deployed. However, at a somewhat lower level of abstraction, the potential risks associated with AI usually appear to be falling into either of two dimensions: 'safety risks' and 'fundamental rights risks'.[16] These two types of risks are just the downside of our expectations of AI and of the promises made by those developing and deploying the technology, that is, that AI will both help by improving health and saving lives and the climate, and assist us in making better decisions, enhancing fairness, and developing into a better society (Figure 12.1).

1. *Traditional (Physical) Safety Risks*

Traditionally, death, personal injury, and damage to property have played a special role within safety and liability frameworks. These traditional types of risks can more specifically be described as 'physical' safety risks, but are normally referred to simply as 'safety risks'. These risks continue to play their very special role in the digital era, but the concept must be understood more broadly to include not only death, personal injury, and damage to property in the traditional sense, but

[16] In previous publications, I have referred to the two types as 'physical' and 'social' risks, see e.g. JP Schneider and C Wendehorst, 'Response to the Public Consultation on the White Paper: On Artificial Intelligence: A European Approach to Excellence and Trust, COM(2020) 65 final' (ELI 2020); C Wendehorst and Y Duller, *Safety and Liability Related Aspects to Software* (*European Commission*, 2021) (hereafter Wendehorst and Duller, 'Safety and Liability') 26 *et seq*; C Wendehorst, 'Strict Liability for AI and Other Emerging Technologies' (2020) JETL (hereafter Wendehorst, 'Strict Liability') 150, 161 *et seq*.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

also damage to data and to the functioning of other digital systems. Where, for example, the malfunctioning of software causes the erasure of important customer data stored by the data holder in some cloud space, this should have the same legal effect as the destruction of a hard disk drive or of paper files with customer data (which is not to say that all data should automatically be treated in exactly the same way as tangible property in the tort liability context).[17] Likewise, where tax management software causes the victim's customer management software to collapse, this must be considered a safety risk, irrespective of whether the customer management software was run on the victim's hard disk drive or somewhere in the cloud within a SaaS scheme. While this is unfortunately still disputed under national tort law,[18] any attempt to draw a line between data stored on a physical medium owned by the victim and data stored otherwise seems to be completely outdated and fails to recognise the functional equivalence of different forms of storage.

2. *Fundamental Rights Risks*

'Fundamental rights risks' are associated with the social dimension of AI. They include discrimination, exploitation, manipulation, humiliation, oppression, and similar undesired effects that are – at least primarily – non-economic (non-material) in nature and that are not just the result of physical harm (as the latter would be dealt with under traditional regimes of compensation for pain and suffering, etc). Such risks have traditionally been dealt with primarily by special legal regimes, such as data protection law, anti-discrimination law or, more recently, law against hate speech on the Internet and similar legal regimes.[19] There is also a growing body of tort law that deals specifically with the infringement of personality rights.[20] Even though the concept of 'fundamental rights' is focused on individual rights, the term 'fundamental rights risks' should be understood more broadly as encompassing also risks of a more collective nature, for example, risks for the rule of law, democracy, and freedom of expression in general.[21]

While the fundamental rights aspect and, therefore, the non-economic aspect of such risks is in the foreground, these risks can, of course, entail economic risks for the affected individual or for society as a whole. For instance, AI systems used for recruitment that favour male applicants create a social risk for female applicants by discriminating against them, but this also leads to adverse economic effects for the affected women.

[17] C Wendehorst, "Liability for Pure Data Loss" in E Karner and others (eds) *Festschrift für Helmut Koziol* (2020) 225 (hereafter Wendehorst, 'Liability for Pure Data Loss').

[18] See Wendehorst, 'Liability for Pure Data Loss' (n 17) 225; G Wagner, '§ 823' in FJ Säcker and others (eds), *Münchener Kommentar zum BGB* (8th ed. 2020) para 245 *et seq*; L Specht, *Konsequenzen der Ökonomisierung informationeller Selbstbestimmung* (2012) 230; F Faust, 'Digitale Wirtschaft: Analoges Recht: Braucht das BGB ein Update?' in Ständige Deputation des Deutschen Juristentages (ed), *Verhandlungen des 71. Deutschen Juristentages – Band I – Gutachten Teil A* (2016), 48.

[19] Regulation (EU) 2016/679, Article 82(1); Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services [2004] OJ L 373/37, Article 8(2); German Network Enforcement Act (Netzwerkdurchsetzungsgesetz, NetzDG, BGBl I S 3352); French Anti-Hate Speech Law (Loi Avia 2020/766); Austrian Anti-Hate Speech Law (Hass-im-Netz-Bekämpfungs-Gesetz, HiNBG, BGBl I 2020/148); Proposal for a Regulation of the European Parliament and the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM (2020) 825 final.

[20] For an overview see G Brüggemeier, AC Ciacchi, and P O'Callaghan, *Personality Rights in European Tort Law* (2010).

[21] C Wendehorst, 'The Proposal for an Artificial Intelligence Act COM(2021) 206 from a Consumer Policy Perspective' (*Federal Ministry Republic of Austria for Social Affairs, Health, Care and Consumer Protection*, 2021) (hereafter Wendehorst, 'The Proposal for an AIA from a Consumer Policy Perspective'), 110.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

3. *Overlaps and In-Between Categories*

The division between safety and fundamental rights risks is generally not always clear-cut and should not be overestimated. There are not only clear overlaps, but also a considerable grey area of a number of important risks. For instance, adverse psychological effects can be a very traditional safety risk,[22] where the effect is a diagnosed illness according to WHO criteria (such as depression), but also a fundamental rights risk that is associated with the social dimension of AI where the effect is not a diagnosed illness, but, for example, just stress or anxiety. It is not always easy to draw a line between the two.[23]

a. Cybersecurity and Similar New Safety Risks

Digitalisation has given rise to a number of very special risks that are not easy to classify. They are essentially safety risks, albeit safety risks of a nature that is somewhat in a grey zone between ‘physical’ and ‘intangible’. Such special safety risks include the ‘data security’ aspect of data protection and privacy (i.e. prevention of data leaks), cybersecurity and harm to the network, and fraud or illegal collusion, to name but a few. They are recognised as relevant safety risks under selected pieces of safety legislation, in particular the Radio Equipment Directive (RED)[24] and the Medical Device Regulation (MDR).[25] Digital risks are also recognised in the Proposal for a Regulation on Machinery Products[26] and the Proposal for a Regulation on General Product Safety,[27] which are intended to replace the Directives currently in force. However, these (digital) risks will often primarily relate to the ‘physical’ dimension of safety, because data theft and manipulation or the breakdown of networks and other essential infrastructures will indirectly, at least in most cases, lead to damage to property in the broader sense or even threaten the health and life of persons.

b. Pure Economic Risks

Pure economic risks[28] are economic risks that are not just the result of the realisation of physical risks, such as personal injury or property damage. Where medical AI causes a surgery to fail, resulting in personal injury and consequently in hospitalisation, the costs of hospitalisation is an economic harm, but not a ‘pure’ economic harm because it results from the personal injury. Where, however, AI manipulates consumers and makes them buy overpriced products, the financial loss caused is not in any way connected with a safety risk and, therefore, qualifies as a pure economic risk (also referred to as immaterial harm). For pure economic risks to be

[22] Article 10:202(1) of the Principles of European Tort Law (hereafter PETL) prepared by the European Group on Tort Law http://egtl.org/PETLEnglish.html.

[23] C van Dam, *European Tort Law* (2006) (hereafter Van Dam, *European Tort Law*) 147.

[24] Article 3(3) Directive 2014/53/EU of the European Parliament and of the Council of 16 April 2014 on the Harmonisation of the Laws of the Member States Relating to the Making Available on the Market of Radio Equipment and Repealing Directive 1999/5/EC [2014] OJ L 153/62.

[25] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, [2017] OJ L 117/1, Annex I, 14.2.

[26] European Commission, ‘Proposal for a Regulation of the European Parliament and of the Council on Machinery Products’ COM (2021) 202 final, Annex III, 1.1.9. and 1.2.1.

[27] European Commission, ‘Proposal for a Regulation of the European Parliament and of the Council on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council, and repealing Council Directive 87/357/EEC and Directive 2001/95/EC of the European Parliament and of the Council’ COM (2021) 346 final, Article 7(1)(h).

[28] PETL, Article 2:102(4); Van Dam, *European Tort Law* (n 20) 169.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

considered legally relevant outside the realm of contractual liability, most legal systems require additional elements, such as fraud or other illegal behaviour or conduct that is considered socially inacceptable.[29] Pure economic risks, at least when legally relevant, might, therefore, be closer to fundamental rights risks.

III. AI AS A CHALLENGE TO EXISTING LIABILITY REGIMES

1. *Classification of Liability Regimes*

While extra-contractual liability law has – beyond product liability law and some few specific areas – so far largely been a matter for the Member States, and while there exists a broad variety of different liability regimes at national level, it is still possible to group liability regimes according to their general characteristics.

a. Fault Liability

Fault liability has been the most important pillar of extra-contractual liability in a majority of European jurisdictions.[30] Liability always requires a sufficient justification for shifting loss from the person who originally suffered the damage (the victim) to a person who caused the damage (the tortfeasor). In the case of fault liability, the fault of the tortfeasor, which is usually either intent or negligence with many different shades and gradations, such as gross negligence or recklessness, is the justification. If damage is caused by mere negligence, further conditions must usually be met, otherwise liability could potentially escalate indefinitely. Jurisdictions use different tools in order to keep liability within reasonable boundaries. Often, there is a requirement that the potential tortfeasor's conduct was somehow objectionable, that is, that it was either violating the law, or public policy, or infringing rights and legally protected interests whose absolute integrity is so vital that any kind of infringement must, per se, be considered as presumably unlawful. The latter is usually the case where human life, health, or bodily integrity are at stake or where the infringement concerns clearly defined property rights.[31]

b. Non-Compliance Liability

Liability may also be triggered by the infringement of particular laws or particular standards whose purpose includes the prevention of harm of the type at hand. We find this type of liability regime both at EU level and at national level. An example for non-compliance liability at EU level is Article 82 of the General Data Protection Regulation (GDPR),[32] which attaches liability to any infringement of the requirements set out by the GDPR. Further, yet very different, examples can be found in EU non-discrimination legislation such as Council Directive 2004/113/EC.[33] Non-discrimination law obliges Member States to introduce into their national legal systems the legal measures necessary to ensure real and effective compensation for loss and

[29] G Brüggemeier, *Tort Law in the European Union* (2nd ed. 2018) para 385; B Wininger and others (eds), *Digest of European Tort Law Volume 2: Essential Cases on Damage* (2011) 383 *et seq*.

[30] For a comparative report, see P Widmer (ed), *Unification of Tort Law: Fault* (2005).

[31] PETL, Article 2:102.

[32] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1.

[33] Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

damage sustained by a person injured as a result of discrimination, in a way which is dissuasive and proportionate to the damage suffered. In this context, Member States must ensure that, when a plaintiff establishes facts from which it may be presumed that there has been direct or indirect discrimination, it shall be for the respondent to prove that there has been no breach of anti-discrimination law.[34] Another example of non-compliance liability can be found in the financial sector. Where issuers of a financial instrument do not publicly disclose inside information concerning them, they become liable for any damage caused by the failure to do so.[35]

At the national level, there may be both general clauses attaching liability to the infringement of protective statutory provisions[36] and specific liability regimes attaching liability to non-compliance with very particular standards. Non-compliance liability is always of an accessory nature, in other words, there needs to be a basic regime setting out in some detail the duties and obligations to be met in order to be considered compliant. It should also be noted that, under a number of national jurisdictions, efforts are being made to impose non-compliance liability only in cases where the potential tortfeasor was at fault.[37]

c. Defect and Mal-Performance Liability

A number of different liability regimes in jurisdictions in Europe may be described as types of 'defect liability' (or, in the case of services, 'mal-performance liability'), although this is certainly not a common technical term. In the extra-contractual realm, the most important form of defect liability is product liability, which has been harmonised by the Product Liability Directive (PLD).[38] Product liability does not require fault on the part of the producer, but it still requires a particular shortcoming in the producer's sphere, in that it requires that the product put into circulation was defective at the time when it left that sphere. The development risk defence (i.e. the defence relying on the fact that the defect, according to the state of the art in science and technology, could not have been detected when the product was put into circulation), which Member States were free to implement or not, moves product liability somewhat into the vicinity of fault liability.[39]

Product liability is only the most conspicuous form of defect liability and the one where the term 'defect' is in fact used. However, when looking more closely at liability regimes in national jurisdictions, it becomes apparent that there is a panoply of different forms of liability that are all based on the unsafe or otherwise objectionable state of a particular object within the liable person's sphere of control. Many of these forms of liability are somewhat at the borderline between fault liability and defect liability, as they are based on a presumption of fault, which the liable person is free to rebut under particular circumstances. Even some forms of vicarious liability under national law may be qualified, at a closer look, as forms of defect or mal-performance liability. For example, vicarious liability may be based on the generally 'unfit' nature of the

[34] Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation [2005] OJ L 204/23, Article 18; Council Directive 2004/113/EC, Article 9; Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation [2000] OJ L 303/16, Article 10.

[35] Explicitly in sections 97 and 98 of the German Securities Trading Act.

[36] See, for example, section 823(2) of the German Civil Code (Bürgerliches Gesetzbuch, BGB) and section 1311 of the Austrian Civil Code (Allgemeines Bürgerliches Gesetzbuch, ABGB).

[37] J Fedtke and U Magnus in BA Koch and H Koziol (eds), Unification of Tort Law: Strict Liability (2002) 147.

[38] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [1985] OJ L 210/29; see for the implementation of the Directive in the Member States WH van Boom and others, 'Product Liability in Europe' in H Koziol and others (eds), *Product Liability Fundamental Questions in a Comparative Perspective* (2017) 255 *et seq*.

[39] NTF Expert Group (n 10) 27 *et seq*.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

relevant auxiliary in terms of personality or skills,[40] or on the fact that the human auxiliary failed to meet a particular objective standard of care.

d. Strict Liability

The term 'strict liability', although often used with a broader meaning, should be reserved for such forms of liability that do not require any kind of defect or mal-performance but are more or less based exclusively on causation. At a closer look, some further requirements beyond causation may have to be met, such as that the risk that ultimately materialised was within the range of risks covered by the relevant liability regime, and there may possibly be defences, such as a *force majeure* defence.[41]

Strict liability is usually imposed only in situations where significant and/or frequent harm may occur despite the absence of any fault or any identifiable defect, mal-performance, or other non-compliance. It is also imposed where such elements would be so difficult for the victim to prove that requiring such proof would lead to massive under-compensation or inefficiency. Paradigm cases are the operation of aircraft, railways, ships, or motor vehicles, although solutions in the EU Member States differ, as does the attitude towards a 'general clause' of strict liability for unforeseen but parallel cases.[42] While there are also examples in national law where something close to strict liability is extended to all objects,[43] this is more or less exceptional and often narrowed down by case law.

2. *Challenges Posed by AI*

The mass rollout of AI and related technologies poses numerous challenges to existing liability regimes. Some of these challenges have their origin in interconnectedness, which is not strictly related to AI, but to digital ecosystems more generally. Other challenges are truly specific to AI.

a. Liability for the Materialisation of Safety Risks

(I) 'COMPLEXITY', 'OPENNESS', AND 'VULNERABILITY' OF DIGITAL ECOSYSTEMS With enhanced connectivity and data flows in the Internet of Things (IoT), everything potentially affects the behaviour of everything, and it may become close to impossible for a victim to prove what exactly caused the damage ('complexity'[44]). For example, where a smart watering system for the garden floods the premises, this may be the effect of the watering system itself being unsafe, but there might also have been an issue with a humidity sensor bought separately, or with the weather data supplied by another provider.

'Openness'[45] means the fact that components are not static but dynamic and are subject to frequent or even continuous change. Products change their safety-relevant features after the product has been put into circulation, for example through the online provision of updates as well as through a variety of different data feeds and cloud-based digital services. This, in fact,

[40] See e.g. section 1315 of the Austrian Civil Code (ABGB).

[41] PETL, Article 7:102(1 a) and Article 5:101 (1); BA Koch and H Koziol, 'Country Report Austria' in BA Koch and H Koziol (eds), *Unification of Tort Law: Strict Liability* (2002) 12, 15, 19.

[42] BA Koch and H Koziol, 'Comparative Conclusions' in BA Koch and H Koziol (eds), *Unification of Tort Law: Strict Liability* (2002) 395 *et seq*.

[43] Responsabilité du fait des choses, Article 1242 Code civil.

[44] NTF Expert Group (n 10) Key Finding no 1(a) 32 *et seq*.

[45] NTF Expert Group (n 10) Key Finding no 1(c) 32 *et seq*.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

means that a victim may not get compensation under liability regimes such as the PLD which exclusively refer to the point in time when a product was first put into circulation.[46]

Connectivity also gives rise to increased 'vulnerability',[47] due to cyber security risks and privacy risks as well as a number of related risks, such as risks of fraud. However, as has been demonstrated by the short survey of existing liability regimes, such risks are not necessarily covered by liability because of a general focus on risks of a 'physical' nature such as death, personal injury, or property damage.

(II) 'AUTONOMY' AND 'OPACITY' AI adds further challenges to an already challenging picture through the features of 'autonomy' and 'opacity'. The term 'autonomy', whose use with regard to machines has often been criticised because of its inextricable link with the free human will, refers to a certain lack of predictability as far as the reaction of the software to unseen instances is concerned. It is in particular when coding of the software has occurred wholly or partially with the help of machine learning[48] that it is difficult to predict how the software will react to each and every situation in the future.[49]

While unpredicted behaviour in new situations nobody had ever thought about may also occur with software of a traditional kind, algorithms created with the help of machine learning cannot easily be analysed, especially not when sophisticated methods of deep learning have been used. This 'opacity' of the code[50] ('black box effect') means that it is not easy to explain why an AI behaved in a particular manner in a given situation, and even less easy to trace that behaviour back to any feature which could be called a 'defect' of the code or to any shortcoming in the development process.

Both autonomy and opacity make it difficult to trace harm back to any kind of intent or negligence on the part of a human actor, which is why fault liability is not an ideal response to risks posed by AI. However, it is also clear that emerging digital technologies, notably AI, make it increasingly difficult to identify a defect due to the autonomy of software and software-driven devices as well as the opacity of the code, which means that defect liability may not be a wholly satisfactory response either.

(III) STRICT AND VICARIOUS LIABILITY AS POSSIBLE RESPONSES As the 'autonomy' and 'opacity' of AI may give rise to exactly the kind of difficulties strict liability is designed to overcome,[51] the further extension of strict liability to AI applications is increasingly being discussed. This would, at the same time, solve some of the problems associated with 'complexity', 'openness', and 'vulnerability' that come with the IoT. For instance, where it is unclear whether the flooding of the premises was due to a defect of the watering system itself, a humidity sensor, or a data feed, it is still clear that the water itself came from the pipes. Thus, if the

[46] Council Directive 85/374/EEC, Article 6(1)(c), Article 7(b); P Machnikowski, 'Conclusions' in P Machnikowski (ed), *European Product Liability: An Analysis of the State of the Art in the Era of New Technologies* (2016) 669, 695.

[47] NTF Expert Group (n 10) Key Finding no 1(g) 32 *et seq*.

[48] Article 3(a) of the EP Resolution on a Civil Liability Regime for AI (n 15) defines 'AI-system' as 'a system that is either software-based or embedded in hardware devices, and that displays behaviour simulating intelligence by, inter alia, collecting and processing data, analysing and interpreting its environment, and by taking action, with some degree of autonomy, to achieve specific goals'.

[49] NTF Expert Group (n 10) Key Finding nos 1(d) and (e) at 32, 33.

[50] NTF Expert Group (n 10) Key Finding no 1(b) 32, 33.

[51] See also NTF Expert Group (n 10) Key Finding no 9, 39 *et seq*.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

legislator introduced strict liability for smart watering systems, this would mean that whoever is the addressee of this strict liability (e.g. the operator or the producer of the watering system) would have to compensate victims for harm suffered from water spread by the system. There have been extensive discussions as to who is the right addressee of liability, and as to which types of risks should ultimately be covered.[52]

Similar effects may be achieved by extending vicarious liability to situations where sophisticated machines are used in lieu of human auxiliaries. Otherwise, parties could escape liability by outsourcing a particular task to a machine rather than to a human auxiliary.[53]

For some time, there has been a debate whether to recognise that highly sophisticated robots, and software agents may themselves be the addressees of liability. The idea of 'electronic personhood' was fuelled by a 2017 European Parliament resolution,[54] but the proposal was met with a great deal of resistance since.[55] Some of the resistance had its roots in ethical considerations,[56] but there are also practical flaws. Being the addressee of liability, AI systems would have to be equipped with funds or with equivalent insurance, which means that electronic personhood is more an additional complication than a solution.[57] Another radical solution proposed is that of replacing liability schemes altogether by insurance or funds so that those suffering harm from AI would be compensated by a general compensation scheme to which, in particular, producers and maybe professional users would be contributing.[58] However, it is meanwhile broadly accepted that such schemes could realistically only be implemented for very particular applications and fields, such as connected driving, but not across the board for a general purpose technology such as AI.[59]

b. Liability for the Materialisation of Fundamental Rights Risks

The main challenge to existing liability schemes is the fact that they are entirely inadequate to address the challenges posed by AI, due to their focus on safety risks. Where fundamental rights risks posed by AI materialise, there is often no fault on the part of those deploying the AI, and it may be close to impossible for a victim to prove that there was fault on the part of the producer. Defect liability, at least as it currently exists under the PLD and under national legal regimes, is entirely focussed on traditional safety risks. This holds true to an even greater extent for strict liability, which, for the time being, is almost exclusively restricted to physical risks. Further, extending vicarious liability to situations where sophisticated machines are deployed in lieu of

[52] Wendehorst and Duller, 'Safety and Liability' (n 16) 93 *et seq*; Wendehorst, 'Strict Liability'(n 16) 165 *et seq*.

[53] NTF Expert Group (n 10) Key Findings nos 18 and 19, 45 *et seq*; H Zech, 'Entscheidungen digitaler autonomer Systeme: Empfehlen sich Regelungen zu Verantwortung und Haftung?' in Ständige Deputation des Deutschen Juristentages (ed), *Verhandlungen des 73. Deutschen Juristentages – Band I – Gutachten Teil A* (2020) (hereafter Zech, 'Entscheidungen digitaler autonomer Systeme') 76 *et seq*.

[54] EP Resolution on Civil Law Rules on Robotics (n 13).

[55] See e.g. the Open Letter to the European Commission Artificial Intelligence and Robotics (2018) www.robotics-openletter.eu/.

[56] Data Ethics Commission, Opinion of the German Data Ethics Commission (*BMJV*, 2019) 219 www.bmjv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html.

[57] NTF Expert Group (n 10) Key Finding no 8, 36 *et seq*.

[58] EP Resolution on Civil Law Rules on Robotics (n 13), paras 57, 59; G Borges, 'New Liability Concepts: The Potential of Insurance and Compensation Funds' in S Lohsse, R Schulze, and D Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (2019) 148 *et seq*; Zech, 'Entscheidungen digitaler autonomer Systeme' (n 53) 105 *et seq*.

[59] J Hanisch, 'Zivilrechtliche Haftungskonzepte für Robotik' in E Hilgendorf (ed), *Robotik im Kontext von Recht und Moral* (2014) 43; J Eichelberger, 'Zivilrechtliche Haftung für KI und Smarte Robotik' in M Ebers and others (eds), *Künstliche Intelligenz und Robotik* (2020) 198.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

human auxiliaries[60] may help also with regard to fundamental rights risks, as long as there is a basis for liability of the hypothetical human auxiliary. Non-compliance liability might possibly be an option, but beyond non-discrimination law, the GDPR, and unfair commercial practices law there is currently not much of a general compliance regime that could serve as a 'backbone' for AI liability. Of course, this 'backbone' could theoretically be created by the emerging AI safety legislation. This is why it is essential to analyse this legislation.

IV. THE EMERGING LANDSCAPE OF AI SAFETY LEGISLATION

While the debate on challenges posed by AI to existing liability regimes is still ongoing, the landscape of AI-relevant product safety law is already changing rapidly, as illustrated by the proposals for a new Machinery Regulation and for the AIA. It is important to understand the emerging safety regimes, because it is only against their background that liability regimes specifically tailored to AI can be properly designed.

1. *The Proposed Machinery Regulation*

a. General Aims and Objectives

The proposed Machinery Regulation aims at modernising the existing machinery safety regime harmonised by the Machinery Directive,[61] in particular with regard to new technologies. This concerns potential risks that originate from a direct human-robot collaboration, risks originating from connected machinery, the phenomenon that software updates affect the 'behaviour' of the machinery after its placing on the market, and the problems associated with risk assessment on machine learning applications before the product is placed on the market. Also, the current regime harmonised by the Machinery Directive still foresees a driver or an operator responsible for the movement of a machine, but fails to set up requirements for autonomous machines. Needless to say, there were also developments to consider and inconsistencies to fix that were not directly related to software and AI. The current list of high-risk machines in Annex I to the Directive was elaborated 15 years ago and is urgently in need of an update.

b. Qualification As High-Risk Machinery

Within the product safety framework for machinery, the qualification of machinery products as high-risk machinery plays an important role. Amongst others, in Annex I, all software ensuring safety functions, including AI systems, and all machinery embedding AI systems ensuring safety functions has been added to the list of high-risk machinery.[62] The fact that all safety components that are software components, and all machinery embedding AI for the purpose of ensuring safety functions, are now included in the list of high-risk machinery automatically means under the proposed Machinery Regulation that, for this kind of machinery, only third party certification will be accepted, even when manufacturers apply the relevant harmonised standards.

A machinery product is included in the list of high-risk machinery products if it poses a particular risk to human health. The notion of 'safety' therefore seems to refer exclusively to risks

[60] NTF Expert Group (n 10) Key Findings nos 18 and 19, 45 *et seq*.

[61] Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC [2006] OJ L 157/24.

[62] Annex I to COM (2021) 202 final, nos 24 and 25.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

of a physical nature. The risk posed by a certain machinery product is, according to Article 5(3) of the Proposal, established based on the combination of the probability of occurrence of harm and the severity of that harm. Factors to be considered in determining the probability and severity of harm include the degree to which each affected person would be impacted by the harm, the number of persons potentially affected, the degree of reversibility of the harm, and indications of harm that have been caused in the past by machinery products which have been used for relevant purposes. However, there are also factors that go more in the direction of 'fundamental rights risks', such as the degree to which potentially affected parties are dependent on the outcome produced by the machinery product, and the degree to which potentially affected parties are in a vulnerable position vis-à-vis the user of the machinery product.

c. Essential Health and Safety Requirements

The essential health and safety requirements that must be met for conformity of high-risk machinery are listed in Annex III. Where machinery uses AI for safety functions, the conformity assessment must consider hazards that may be generated during the lifecycle of the machinery as an intended evolution of its fully or partially evolving behaviour or logic.[63] As far as human-machine collaboration is concerned, a machinery product with fully or partially evolving behaviour or logic that is designed to operate with varying levels of autonomy must be adapted to respond to people adequately and appropriately; this must occur verbally through words or nonverbally through gestures, facial expressions, or body movement. It must also communicate its planned actions (what it is going to do and why) to operators in a comprehensible manner.[64]

Largely, however, AI-specific aspects are referred to in the future AIA, that is, where the machinery product integrates an AI system, the machinery risk assessment must consider the risk assessment for that AI system that has been carried out pursuant to the AIA.[65]

2. *The Proposed Artificial Intelligence Act*

a. General Aims and Objectives

The AIA Proposal of 21 April 2021 aims at ensuring that AI systems placed on the Union market and used in the Union are safe and respect existing law on fundamental rights and Union values, and at enhancing governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems. At the same time, efforts are being made to ensure legal certainty in order to facilitate investment and innovation in AI and to facilitate the development of a single market for AI applications and prevent market fragmentation. The AIA is complementary to existing data protection law (in particular the GDPR and the Law Enforcement Directive[66]), non-discrimination law, and consumer protection law.

As regards high-risk AI systems, which are safety components of products, the AIA will be integrated into the existing and future product safety legislation. For high-risk AI systems related

[63] Annex III to COM (2021) 202 final, no 1(c).

[64] Annex III to COM (2021) 202 final, no 1.3.7.

[65] Annex III to COM (2021) 202 final, no 1(c).

[66] Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA [2016] OJ L 119/89.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

to products covered by the New Legislative Framework (NLF) legislation (e.g. machinery, medical devices, toys), the requirements for AI systems set out in the AIA will be checked as part of the existing conformity assessment procedures under the relevant NLF legislation.[67] The latter may, at the same time, include further AI-specific requirements relevant only in a particular sector. AI systems related to products covered by relevant 'old approach' legislation (e.g. aviation, motor vehicles)[68] are not directly covered by the AIA, though.[69]

b. The Risk-Based Approach

The AIA Proposal follows a risk-based approach, differentiating between uses of AI that create an unacceptable risk, a high risk, a limited risk, and a low or minimal risk.

(i) PROHIBITED AI PRACTICES Title II lists some narrowly defined AI systems whose use is considered unacceptable as contravening EU values and violating fundamental rights, such as manipulation through subliminal techniques or exploitation of group-specific vulnerabilities (e.g. children) in a manner that is likely to cause affected persons psychological or physical harm. The Proposal also prohibits general-purpose social scoring by public authorities and, subject to a range of exceptions, the use of 'real time' remote biometric identification systems in publicly accessible spaces for law enforcement purposes.[70]

(ii) HIGH-RISK AI SYSTEMS Title III contains mandatory essential requirements for AI systems qualified as 'high-risk' AI systems, defined as systems that create a high risk to the health and safety or fundamental rights of natural persons. There are two main categories of high-risk AI systems: AI systems used as a safety component of products that are subject to third party *ex ante* conformity assessment under NLF legislation listed in Annex II; and other stand-alone AI systems explicitly listed in Annex III. The systems listed in Annex III, as it currently stands, more or less exclusively address fundamental rights risks. This includes biometric identification and categorisation of natural persons; education and vocational training; employment; workers management and access to self-employment; access to, and enjoyment of, essential private services, public services, and benefits; law enforcement; migration, asylum and border control management; and administration of justice and democratic processes. The only exception is the 'management and operation of critical infrastructure'[71] as the latter poses a systemic risk of a more physical nature rather than a fundamental rights risk.

The Commission may, from time to time, expand the list of high-risk AI systems used within certain pre-defined areas, by applying a set of criteria and risk assessment methodology. The risk assessment criteria listed in Article 7(2) are similar to those listed in the relevant Article of the proposed Machinery Regulation,[72] with two main exceptions: Reference is not only made to risks for the health of persons, but also to risks for the 'health and safety or … fundamental rights'. Also, an additional criterion to consider is the extent to which existing Union legislation already provides for effective measures of redress in relation to the risks posed by an AI system (with the exclusion of claims for damages) and the existence of effective measures to prevent or

[67] Article 6(1)(b); Recital 63 COM (2021) 202 final.

[68] Annex II section B to COM (2021) 202 final.

[69] Article 2(2)(2) COM (2021) 202 final.

[70] Wendehorst, 'The Proposal for an AIA from a Consumer Policy Perspective' (n 21) 75.

[71] Annex III to COM (2021) 202 final, no 2.

[72] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on Machinery Products' COM (2021) 202 final, Article 5(3).

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

substantially minimise those risks. For the purpose of future classification of additional AI systems as 'high-risk' systems, safety risks and fundamental rights risks are treated in the same manner and are not dealt with separately.

(III) AI SYSTEMS SUBJECT TO SPECIFIC TRANSPARENCY OBLIGATIONS Title IV is devoted to AI systems that are subject to enhanced transparency obligations. This concerns, for example, AI systems that may be mistaken for human actors, deep fakes, emotion recognition systems, and biometric categorisation systems.[73] It is important to note, though, that Titles III and IV are not mutually exclusive, i.e. an AI system that qualifies as a 'high-risk' system for the purpose of Title III may still fall under IV as well.

c. Legal Requirements and Conformity Assessment for High-Risk AI Systems

Legal requirements set out in Title III for high-risk AI systems address data and data governance, documentation and record keeping, transparency and provision of information to users, human oversight, robustness, accuracy, and security. By and large, and with regard to the AI system, the same requirements apply irrespective of whether what is at stake is the safety component of a toy robot or a connected household device falling under the RED, or an AI system intended to be used for the selection and evaluation of applicants in the course of a recruitment procedure. This may not be particularly convincing, because the safety requirements with regard to the toy robot or the connected household device are very different to the safety requirements with regard to the recruitment software. However, due to the general nature of the requirements and obligations listed in the Proposal, it may still be the better choice to deal with the two risk categories under identical provisions.

Obligations with regard to these requirements are largely placed on producers (called 'providers') of high-risk AI systems, but proportionate obligations are also placed on (professional) users and other participants across the AI value chain (such as importers, distributors, and authorised representatives) consistent with other modern product safety legislation. The Proposal sets out a framework for notified bodies to be involved as independent third parties in conformity assessment procedures. AI systems used as safety components of products regulated under the NLF, such as machinery or toys, are subject to the same compliance and enforcement mechanisms of the products of which they are a component, but in the course of applying these mechanisms the requirements imposed by the AIA must be ensured as well. New *ex ante* re-assessments of the conformity will be needed in case of substantial modifications to the AI systems.

As regards stand-alone high-risk AI systems, which are currently not covered by product safety legislation, a new compliance and enforcement mechanism is established along the lines of existing NLF legislation. However, with the exception of remote biometric identification systems, such high-risk AI systems are only subject to self-assessment of conformity by the providers. The justification provided in the explanatory notes[74] is that the combination with strong *ex post* enforcement would be an effective and reasonable solution, given the early phase of the regulatory intervention and the fact the AI sector is very innovative and expertise for auditing is only now being accumulated.[75]

[73] Wendehorst, 'The Proposal for an AIA from a Consumer Policy Perspective' (n 22) 27; C Wendehorst and Y Duller, 'Biometric Recognition and Behavioral Detection' (*European Parlament*, 2021), 63; C Wendehorst and J Hirtenlehner, 'Outlook on the future regulatory requirements for AI in Europe' (2022), 35.

[74] COM (2021) 206 final, explanatory note no 64.

[75] Critical T Schmidt and S Voeneky, Chapter 8, in this volume.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

V. THE EMERGING LANDSCAPE OF AI LIABILITY LEGISLATION

While Commission proposals on AI liability, which were initially planned for the first quarter of 2022, have meanwhile been postponed to the third quarter of 2022, a draft Regulation by the European Parliament has been on the table since October 2020.[76] It was prepared in parallel with the Commission's White Paper on AI and the preparatory work for the AIA Proposal and has clearly been influenced by work at Commission level.

1. *The European Parliament's Proposal for a Regulation on AI Liability*

The cornerstone of the EP Proposal for the regulation of AI liability is a strict liability regime for the operators of 'high-risk' AI systems enumeratively listed in an Annex, accompanied by an enhanced regime of fault liability for the operators of other AI systems.

a. Strict Operator Liability for High-Risk AI Systems

According to Article 4 of the EP Proposal, operators of AI systems shall be strictly liable for any harm or damage that was caused by a physical or virtual activity, device, or process driven by an AI system. The EP Proposal ultimately adopted the division into 'frontend operator' (i.e. the person deploying the AI system) and 'backend operator' (i.e. the person that continuously controls safety-relevant features of the AI system, such as by providing updates or cloud services) that had been developed by the author of this paper and included in the 2019 EG-NTF report.[77] According to the final version of the EP Proposal, not only the frontend operator, but also the backend operator may become strictly liable. However, the backend operator's liability is covered only if it is not already covered by the PLD.[78] The only defence available to the operator is *force majeure*.[79] For the AI systems subject to strict liability, mandatory insurance is being proposed.[80]

'High-risk' AI systems for the purpose of the proposed Regulation are to be exhaustively listed in an Annex. Interestingly, the final version of the Proposal was published with the Annex left blank. The Annex attached to the first published draft from April 2020 had met with heavy resistance due to its many inconsistencies, and it may have proved too difficult to agree on a better version. Also, it seemed opportune to wait for the list of 'high-risk' AI applications that would be attached to the AIA. In any case, given the rapid technological developments and the required technical expertise, the idea is that the Commission should review the Annex without undue delay, but at least every six months, and if necessary, amend it through a delegated act.[81]

b. Enhanced Fault Liability for Other AI Systems

The EP Proposal does not only include a strict liability regime for 'high-risk' applications, but also a harmonised regime of rather strictish fault liability for all other AI systems. Article 8 provides for fault-based liability for 'any harm or damage that was caused by a physical or virtual activity, device or process driven by the AI-system', and fault is presumed (i.e. it is for the operator to show that the harm or damage was caused without his or her fault).[82] In doing so, the

[76] EP Resolution on a Civil Liability Regime for AI (n 15).

[77] NTF Expert Group (n 10) Key Findings nos 10 and 11.

[78] See Article 3(e).

[79] See Article 4(3).

[80] *Cf.* EP Resolution on a Civil Liability Regime for AI (n 15) Article 4(4).

[81] EP Resolution on a Civil Liability Regime for AI (n 15) Recommendation to the Commission no 16.

[82] In fact, the drafting is not very clear with regard to this point. Recital 17 seems to underline that fault is always presumed and that the operators need to exonerate themselves. However, Recital 19 also refers to proof of fault by the victim.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

operator may rely on either of the following grounds: The first ground is that the AI-system was activated without his or her knowledge while all reasonable and necessary measures to avoid such activation outside of the operator's control were taken. The second ground is that due diligence was observed by performing all the following actions: selecting a suitable AI-system for the right task and skills, putting the AI-system duly into operation, monitoring the activities, and maintaining the operational reliability by regularly installing all available updates. It looks as if these two grounds are the only grounds by means of which operators can exonerate themselves, but Recital 18 also allows for a different interpretation, namely, that the two options listed in Article 8(2) should just facilitate exoneration by establishing 'counter-presumptions'.

The proposed fault liability regime is problematic not only because of the lack of clarity in drafting, but also because Article 8(2)(b) might be unreasonably strict, as it seems that the operator must demonstrate due diligence in all aspects mentioned, even if it is clear that lack of an update cannot have caused the damage. More importantly, in the absence of any restriction to professional operators, even consumers would face this type of enhanced liability for any kind of AI device, from a smart lawnmower to a smart kitchen stove. This would mean burdening consumers with obligations to ensure that updates are properly installed, irrespective of their concrete digital skills, and possibly confronting them with liability risks they would hardly ever have had to bear under national legal systems.

c. Liability for Physical and Certain Immaterial Harm

Article 2(1) of the Proposal declares the proposed Regulation to apply where an AI system has caused 'harm or damage to the life, health, physical integrity of a natural person, to the property of a natural or legal person or has caused significant immaterial harm resulting in a verifiable economic loss'. Article 3(i) provides for a corresponding definition of 'harm or damage'. While life, health, physical integrity, and property were clearly to be expected in such a legislative framework, the inclusion of 'significant immaterial harm resulting in a verifiable economic loss' came as a surprise. If immaterial harm or the economic consequences resulting from it – such as loss of earnings due to stress and anxiety that do not qualify as a recognised illness – is compensated through a strict liability regime whose only threshold is causation,[84] the situations where compensation is due are potentially endless and difficult to cover by way of insurance.[85]

This is so because there is no general duty not to cause significant immaterial harm of any kind to others, unless it is caused by way of non-compliant conduct (such as by infringing the law or by intentionally acting in a way that is incompatible with public policy). For instance, where AI used for recruitment procedures leads to a recommendation not to employ a particular candidate, and if that candidate, therefore, suffers economic loss by not receiving the job offer, full compensation under the EP Proposal for a Regulation would be due even if the recommendation was absolutely well-founded and if there was no discrimination or other objectionable element involved. While some passages of the report seem to choose somewhat more cautious formulations, calling upon the Commission to conduct further research,[86] Recital 16 explains very firmly that 'significant immaterial harm' should be understood as meaning harm as a result of which the affected person suffers considerable detriment, an objective and demonstrable impairment of his or her personal interests and an economic loss calculated having regard, for example, to annual average figures of past revenues and other relevant circumstances.

[83] EP Resolution on a Civil Liability Regime for AI (n 15) Article 4(1).

[84] *Cf.* T Schmidt and S Voeneky, Chapter 8, in this volume, who suggest that companies that develop or produce high-risk AI should contribute to a fund that covers damages caused by AI-driven high-risk products or services.

[85] EP Resolution on a Civil Liability Regime for AI (n 15) Recommendation to the Commission no 19.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

2. *Can the EP Proposal be Linked to the AIA Proposal?*

The 2020 White Paper on AI, the EP's 2020 Proposal for an AI Liability Regulation, and the 2021 Commission Proposals for an AIA and for a new Machinery Regulation clearly have a number of parallels. They range from some identical terminology (e.g. 'AI system', 'high-risk') to the legislative technique of exhaustively listing 'high-risk' AI systems in an Annex, combined with the option for the European Commission to amend the Annex in a rather flexible procedure through delegated acts. So the question arises whether it would be possible to link an AI liability regime along the lines of the EP Proposal with the AIA Proposal in a way that the legal requirements and obligations perspective matches the liability perspective.

a. Can an AI Liability Regulation Refer to the AIA List of 'High-Risk' Systems?

The first question that arises is whether the list of 'high-risk' AI systems in the AI Liability Regulation can be identical to the list of 'high-risk' AI systems under the AIA. However, as tempting as it may be to simply refer to the AIA, it would lead to overreaching and inappropriate results. The justification for imposing strict liability that the relevant product or activity leads to significant and/or frequent harm despite the absence of any fault or any identifiable defect, mal-performance, or non-compliance does not coincide with the justification for imposing particular precautionary measures against unsafe products. While the AI systems for which strict liability is justified will most likely be a subset of the AI systems for which enhanced safety measures are justified, by far not all AI systems of the latter type should be included in a strict liability regime, for example, when they are normally safe except when clearly defective. This is underlined by the fact that the relevant players are not identical. While safety requirements are primarily addressed at the level of producers ('providers' in the AIA terminology), the EP Proposal suggests imposing strict AI liability primarily on the frontend operators ('users' in the AIA terminology), but also on the backend operators (a concept missing in the AIA). So even if something along the lines of the EP Proposal became the law it would be imperative to draft a liability-specific Annex defining 'high-risk' AI systems specifically for liability purposes. This could, for example, include big AI-driven cleaning or lawnmower robots used in public spaces, but not a small vacuum cleaner or toy robot.

b. Can the AIA Keep Liability for Immaterial Harm within Reasonable Boundaries?

As concerns fundamental rights risks, the current approach taken by the EP Proposal, which considers strict liability (alongside fault liability) for 'significant immaterial harm that results in a verifiable economic loss', has already been discarded earlier in this chapter[86] because of its failure to keep liability within any reasonable boundaries. However, the question arises whether the AIA Proposal can now assist in solving this problem.

One way of attaching liability immediately to the AIA Proposal seems to be attaching liability to the engagement in any prohibited AI practice within the meaning of Title II of the AIA Proposal, which could lead to the compensation of both material and immaterial harm thereby caused. This would be a model of non-compliance liability and fit easily into existing non-discrimination, data protection, and consumer protection legislation, all of which provide for liability for damages where harm has been caused by the engagement in prohibited practices.

Another option would be to restrict liability for immaterial harm to cases of non-conformity with the legal requirements in Title III Chapter 2 of the AIA. For instance, where training, validation, or testing data for recruitment AI fail to be relevant, representative, free of errors, and complete, as

[86] See V 1(c).

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

required by Article 10(4) of the AIA Proposal, the provider could be liable if an applicant was falsely filtered out by the system despite being objectively better qualified. However, it soon transpires that the legal requirements included in Title III Chapter 2 of the AIA Proposal are not optimally suited as a basis for defect liability. For many of the requirements are not so much ends in themselves that would automatically mean an AI system violates fundamental rights. Rather, some of them resemble due diligence standards that must be met during AI development, either as a quality-enhancing measure (e.g. data governance) or to facilitate monitoring (e.g. record-keeping). Non-conformity with such requirements could, therefore, justify a shift of the burden of proof, but should not in itself trigger liability. Thus, in the case of the recruitment AI system, non-conformity of training data with Article 10 should not lead to a final determination of liability but rather to the presumption that the resulting AI was defective.

VI. POSSIBLE PILLARS OF FUTURE AI LIABILITY LAW

If the AIA Proposal as it currently stands is not optimally suited for functioning as a 'backbone' for AI liability, this does not mean that the AIA as such cannot fulfil this function. Upon a closer look, not much would have to be changed in the AIA to make it an appropriate basis for future legal regimes on AI liability. At the end of the day, liability for damages caused by AI systems may have to rest on different pillars, all of which would have to rely on, or at least be aligned with, provisions in the AIA and further product safety and other law.

1. *Product Liability for AI*

The first obvious link between the AIA (and other product safety law) on the one hand and liability law on the other could be established within product liability law, which relies on the PLD. Meanwhile, it is widely accepted that the PLD must in any case be adapted to the challenges of digital ecosystems at large.[87]

a. Traditional Safety Risks

With regard to the reform of the PLD, the debate has so far been focused entirely on safety risks. Already with regard to these risks, the PLD as it currently stands is not fit to meet the challenges posed by digitalisation, not least in the light of uncertainties with regard to its scope (e.g. concerning self-standing software, including AI) and its focus on the point in time when a product is put into circulation, which fails to take into account updates, data feeds, and machine learning.[88] Where AI is involved, a victim may face particular difficulties showing that the AI system was defective. This is why no defect of the AI should have to be established by the victim for AI-specific harm caused by AI-driven products. Rather, it should be sufficient for the victim to prove that the harm was caused by an incident that might have something specifically to do with the AI (e.g. the cleaning robot making a sudden move in the direction of the victim) as contrasted with other incidents (e.g. the victim stumbling over the powered-off cleaning robot).[89]

[87] Among the plethora of pleas made in this direction, see only C Twigg-Flesner in European Law Institute (ELI) (ed), *Guiding Principles for Updating the Product Liability Directive for the Digital Age* (2021) https://europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Guiding_Principles_for_Updating_the_PLD_for_the_Digital_Age.pdf.

[88] Wendehorst and Duller, 'Safety and Liability' (n 16) 68; Koch, 'Product Liability 2.0 – Mere Update or New Version?' (n 14) 102.

[89] Wendehorst and Duller, 'Safety and Liability' (n 16) 6, 93.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

b. Product Liability for Products Falling Short of 'Fundamental Rights Safety'?

As has been pointed out, the AIA Proposal also addresses fundamental rights risks. This raises the question whether also product liability might, in the future, include liability for products with a 'fundamental rights defect' or falling short of 'fundamental rights safety'.

The legal requirements described in Title III Chapter 2 of the AIA Proposal address some cloudy notion of 'adverse impact on the fundamental rights' of persons, including non-discrimination and gender equality, data protection and privacy, and the rights of the child. However, they fail to state – either in a positive or in a negative manner – what exactly the legal requirements are designed to achieve or to prevent. It is rather obvious that discrimination as far as prohibited by EU non-discrimination law, or data processing as far as prohibited by EU data protection law, is among the core effects to be prevented. However, given the much more 'fuzzy' nature of fundamental rights risks as compared with traditional safety risks, and given that there is a floating spectrum of beneficial or adverse impact on a broad variety of different fundamental rights, it is very difficult to impose liability for the materialisation of fundamental rights risks as such.

In order to achieve liability for the materialisation of fundamental rights risks as such, the first step must be to formulate an equivalent to the established concept of 'safety' in traditional product safety legislation. As far as traditional safety risks are concerned, it is possible for Article 6 (1) of the PLD to simply state: 'A product is defective when it does not provide the safety which a person is entitled to expect, taking all circumstances into account [. . .]', implicitly referring to the bulk of existing product safety law that is designed to protect 'the safety and health of persons' and similar traditional notions of safety. A corresponding concept of 'fundamental rights safety' could theoretically be derived from the AIA, in particular from the requirements for high-risk AI systems listed in Chapter 2 of Title III of the current proposal. However, in order to make these requirements operational for purposes of liability law they would have to be divided into two groups. Requirements which constitute 'AI-specific safety' (which would, by and large, be the requirements listed in Articles 13 through 15 of the draft AIA) would have to be seen as clearly separated from the requirements that are about managing safety (mostly Article 9), increasing the likelihood of safety (selected aspects of which are listed in Article 10), or documenting safety (Articles 11 and 12). Shortcomings in the technical documentation or in logging capabilities, for instance, should not be seen as a lack of 'fundamental rights safety' as such, but should rather trigger proof-related consequences in the liability context. Where technical documentation or logging capabilities are missing, or where the producer withholds logging data that would be available and potentially relevant, there could be a presumption that the missing information would have been to the detriment of the producer. Where, on the other hand, an AI system is not as accurate and robust as stated in its description or as could reasonably be expected from an AI system of the relevant kind, and therefore harm occurs (e.g. recruitment software assessing candidates has a strong gender bias and therefore female applicants are discriminated against), this lack of accuracy or robustness might trigger liability of the provider under an extended scheme of product liability. Designing such an extended scheme of product liability would, without doubt, remain to be challenging.

2. *Strict Operator Liability for 'High-Physical-Risk' Devices*

As far as death, personal injury, or property damage caused by a 'high-risk' product that includes AI for safety-relevant functions is concerned, strict liability seems to be a proper response. Again, the question arises whether the AIA can be made operational for the purposes of liability law.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

a. Why AI Liability Law Needs to be More Selective than AI Safety Law

As has already been pointed out,[90] not every product that qualifies as a 'high-risk' product under the AIA fulfils the requirements that should be met for justifying strict liability (and the accompanying burden of insurance). For instance, a small robot vacuum cleaner may, under the future Machinery Regulation (if the current draft were enacted as is), be automatically classified as 'high-risk' and be subject to third party conformity assessment. It would, therefore, at least if the AI component fulfils a safety function, automatically be classified as a 'high-risk' AI system also under AIA. Similarly, a toy robot vehicle for children using AI for a safety function would be qualified as 'high-risk' under the AIA in cases where that toy is subject to third party conformity assessment,[91] (e.g. in any case where no harmonised standards exist that cover all safety requirements, or the producer has deviated from the standard).[92]

However, it would arguably be exaggerated to impose strict liability for harm caused by small toy robots or robot vacuum cleaners, in particular if that strict liability is imposed on operators. Those machines hardly ever cause significant physical harm by themselves, and if they do, it is usually because it was improper for the (frontend) operator to deploy them in the particular situation, such as where the operator of a retirement home uses an unsupervised cleaning robot in places and at times when elderly residents might stumble over it. Another possibility is that the machine is defective, for example, the vacuum cleaner, which is normally only used during the night in areas that are locked for residents, suddenly breaks loose and starts hovering when elderly residents are leaving the dining room. The problem is not so much that it would be inappropriate in the case of the retirement home to make its operator strictly liable for damage caused by the cleaning robot. Rather, the problem is that if all operators of small vacuum cleaner robots (including the millions of businesses that use them for cleaning their office space during the night, or even consumers) had to face strict liability and had to take out corresponding insurance, this would be extremely inefficient and benefit no one but the insurance industry.

b. Differentiating 'High-Risk' and 'High-Physical-Risk-As-Such'

The AIA could, therefore, be made fully operational as a 'backbone' to AI liability law if its Article 6 with Annex II drew a distinction between AI systems that are – for whatever inner logic the relevant sectoral NLF product safety legislation may follow – subject to third party conformity assessment, and AI systems that create a high physical risk as such. Needless to say, the two groups would not be mutually exclusive, as AI systems that create a high physical risk as such will often be subject to third party conformity assessments under the relevant product safety law. On the other hand, it will often be AI systems governed by 'old approach' legislation[93] that pose a high physical risk to the safety of persons as such. This means that the AIA could provide a better basis for AI liability law if these two groups of AI systems could be separated and better differentiated, either by way of restructuring and slightly redrafting Article 6 and Annex II or by drawing that distinction in a separate legal instrument on AI liability.

[90] See sub V 2(a).

[91] Article 19(3) of Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys [2009] OJ L 170/1 1, last amended by Commission Directive (EU) 2018/725 of 16 May 2018.

[92] As set out in Article 10 and Annex II of Directive 2009/48/EC. Note that the requirements are so far focused on mechanical/physical properties (e.g. sharp edges and weight), flammability, chemicals, and heavy metals restrictions, so there will be only very few AI-driven toys qualifying as 'high-risk' under the AIA.

[93] As listed in section B of Annex II and largely exempt from the AIA itself by Article 2(2) COM (2021) 202 final.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

c. Avoiding Inconsistencies with Regard to Human-Driven Devices

However, it should also be borne in mind that strict liability for physical risks caused by AI-driven devices might create significant inconsistencies if not accompanied by strict liability for the same type of devices where those devices are not AI-driven but steered by humans or by technology other than AI. A victim run over by a vehicle does not care that much whether the vehicle was AI-driven or not. So if strict liability is found to be appropriate for a particular type of device of a certain minimum weight running at a certain minimum speed in public spaces (or other spaces where they typically get into contact with persons involved with the operation), this will normally be the case irrespective of whether the device is human-driven or AI-driven. For instance, large cleaning machines, lawnmowers, or delivery vehicles in public spaces might generally have to be included in strict liability regimes even where, in the relevant jurisdiction, this is so far not the case. So a strict liability regime should, at the end of the day, not be restricted to AI systems.

3. *Vicarious Operator Liability*

Vicarious liability in the sense of liability for the acts and omissions of others, such as (human) auxiliaries, might be yet another pillar of future AI liability.

a. The 'Accountability Gap' that Exists in a Variety of Contexts

Part of the problem with existing liability regimes in Member States is associated with the absence, in most legal systems, of vicarious liability for the mal-functioning of machines. Where a human cleaner knocks over a person passing by, or where a human bank clerk miscalculates a customer's credit score, there is usually fault liability of either the human auxiliary that was acting, or their employer, or both. Where, however, the person passing by is knocked over by a cleaning robot, or the credit score miscalculated by credit scoring AI, it is well possible that no one is liable at all. The AI system itself cannot be liable, but its operator may not be liable either if that operator can demonstrate that they have bought the AI system from a recognised provider and complied with all monitoring and similar duties. The producer will often not be liable as a defect in the AI system is sometimes difficult to prove, and in any case product liability (unless it will be significantly extended) only covers personal injury and property damage.

Vicarious liability would be a solution, but the rules on liability for acts or omissions of others differ vastly across the Member States and some courts insist that this kind of liability remains restricted to human auxiliaries.[94] Due to the fact that the application of vicarious liability, either directly or by analogy, is uncertain, an 'accountability gap' may exist, as very harmful activities could be conducted without anyone taking responsibility. This concerns both contexts where fault liability would normally apply and contexts where there would be non-compliance liability, and possibly other contexts.

b. Statutory or Contractual Duty on the Part of the Principal

Vicarious AI liability can only go as far as the operator of the AI would itself be liable, under national law, for violation of the same standard of conduct. This means that there must exist some statutory or contractual duty, in particular a duty of diligence, on the part of the operator. Such duties may exist in a variety of contexts, from professional care to recruitment to credit scoring to pricing, and vicarious liability may become relevant for a variety of legal frameworks, from traditional areas of tort law to non-discrimination law to data protection law to consumer and competition law.

[94] NTF Expert Group (n 10) 24 *et seq*.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

Such duties could also follow from the AIA. It is, in particular, the engagement in prohibited AI practices that should lead to liability, irrespective of whether the operator was acting intentionally or negligently with regard to the fact that, for example, the AI was exploiting age-specific vulnerabilities. With an associated liability scheme in mind, it becomes even more apparent, though, that the very 'pointillistic' style of Title II of the AIA Proposal is a problem and that, if fundamental rights protection is taken seriously, it would have been necessary to have a more complete list of blacklisted AI practices plus ideally a general clause to cover unforeseen cases.

c. A Harmonised Regime of Vicarious Liability

A new European scheme of vicarious liability might restrict itself to ensuring that a principal that employs AI for a sophisticated task faces the same liability under existing Member State law as a principal that employs a human auxiliary.[95] For example, a professional user of an AI system would be liable for harm caused by any lack of accuracy or other shortcomings in the operation of the system to the same extent as that user would be liable (under the applicable national law) for the acts or omissions of a human employee mandated with the same task as the AI system. Where a human would not have been able to fulfil the same task, such as where the task requires computing capabilities exceeding those of humans, the point of reference for determining the required level of performance would be available comparable technology which the user could be expected to use.[96]

However, the EU legislator could also go one step further and introduce a fully harmonised concept of vicarious liability that does not suffer from the outset from the shortcomings we see in existing national concepts. By and large, this new European scheme of vicarious liability could provide that a business or public authority is liable for damage caused by its human auxiliaries acting within the scope of their functions, or any AI employed by the business or public authority, where these auxiliaries or AI fail to perform – for whatever reason – at the standard that could reasonably be expected from them.[97] This comes close to strict liability insofar as it requires neither fault nor a defect (or general lack of reliability in the case of human auxiliaries), but some output that does not meet the standards of conduct to be expected from a business or public authority in the fulfilment of their functions. What this level of quality is, depends on the task to be fulfilled. For instance, if it is about assessing the creditworthiness of a customer seeking credit, it would be the duty to provide proper assessment along the lines of any criteria prescribed by the law or stated by the business, and if it is about assessing candidates for a vacant position, it is again about assessing them properly, without any prohibited discrimination and duly taking into account the qualifications required for the position. Vicarious liability would, in any case, cover both safety risks and fundamental rights risks.

4. *Non-Compliance and Fault Liability*

Last but certainly not least, non-compliance and fault liability can also play an important role in the future landscape of liability for AI. In very much the same manner as Article 82 of the GDPR provides for liability of a controller or processor where that controller or processor violates their obligations under the GDPR, there could be liability under the AIA, or in a separate piece of legislation, where a provider, user or other economic operator covered by the AIA fails to comply with relevant AIA provisions, thereby causing relevant harm. This non-compliance liability

[95] Wendehorst and Duller, 'Safety and Liability' (n 16) 92.

[96] NTF Expert Group (n 10) Key Findings nos 18 and 19.

[97] This would amount to a combination between Article 6:102 (Liability for Auxiliaries) and Article 4:202 (Enterprise Liability) PETL.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press

might complement general fault liabiity that would continue to co-exist as a general baseline for extra-contractual liability. A breach of a duty of care that would constitute negligence could include deploying AI for a task it was not designed for, failing to provide for appropriate human oversight and other safeguards or failing to provide for necessary long-term monitoring and maintenance. Non-compliance liability and fault liability could also be merged, such as by alleviating the burden of proof for the victim under fault liability, or even reversing that burden, where obligations under the AIA have failed to be complied with.

VII. CONCLUSIONS

The potential risks associated with AI appear as normally falling into either of two dimensions: (a) 'safety risks' (i.e. death, personal injury, damage to property etc.) caused by unsafe products and activities involving AI and (b) 'fundamental rights risks' (i.e. discrimination, total surveillance, manipulation, exploitation, etc.), including risks for society at large, caused by inappropriate decisions made with the help of AI or otherwise inappropriate deployment of AI. While safety risks are highly relevant also in the AI context, fundamental rights risks are much more AI-specific.

Existing extra-contractual liability regimes can essentially be divided into four categories: fault liability, non-compliance liability, defect or mal-performance liability, and strict liability in the narrower sense. Vicarious liability can normally also be analysed as falling into one of these categories. Three out of the four categories of liability regimes are either restricted to, or heavily focused on, traditional safety risks such as death, personal injury, or property damage. It is only non-compliance liability, such as can be found in the GDPR or as an annex to EU non-discrimination law or consumer protection law, that frequently addresses also harm resulting from fundamental rights risks. Despite the fact that fundamental rights risks are more AI specific, liability for such risks seems to be largely unchartered territory, and the debate around liability for AI has largely been restricted to safety risks.

At the level of AI safety law, fundamental rights risks are now being addressed by way of prohibiting certain AI practices and by imposing mandatory legal requirements for other 'high-risk' AI systems, such as concerning data and data governance, transparency, and human oversight. While it is not impossible to use the emerging AI safety regime as a 'backbone' for the future AI liability regime, the AIA proposal, as it currently stands, is not optimally suited to help address liability for fundamental rights risks.

The future AI liability law could rest on several different pillars, such as: (a) a revised regime of product liability, which might even include liability for lack of 'fundamental rights safety'; (b) strict operator liability for death, personal injury, property damage, and possibly further safety risks caused by 'high-physical-risk' devices; (c) vicarious operator liability for mal-performance of functions carried out in the course of business activities or activities of a public authority; and (d) fault and/or non-compliance liability for the operator's own negligence and/or failure to comply with obligations following from, in particular, the AIA.

While it would be desirable to have an AI safety regime that allows an AI liability regime to dock on, it becomes apparent that the AIA Proposal has, regrettably, not been drafted with liability law in mind. Further negotiations about the AIA Proposal and the preparatory work on a future AI liability regime as well as on a potential revision of the PLD should, for the sake of consistency of Union law and of legal certainty, be more closely aligned.

https://doi.org/10.1017/9781009207898.016 Published online by Cambridge University Press