Autonomous Artificial Intelligence and Uncontemplated Hazards: Towards the Optimal Regulatory Framework

Mitja KOVAC*

The issue of super-intelligent artificial intelligence (AI) has begun to attract ever more attention in economics, law, sociology and philosophy studies. A new industrial revolution is being unleashed, and it is vital that lawmakers address the systemic challenges it is bringing while regulating its economic and social consequences. This paper sets out recommendations to ensure informed regulatory intervention covering potential uncontemplated AI-related risks. If AI evolves in ways unintended by its designers, the judgment-proof problem of existing legal persons engaged with AI might undermine the deterrence and insurance goals of classic tort law, which consequently might fail to ensure optimal risk internalisation and precaution. This paper also argues that, due to identified shortcomings, the debate on the different approaches to controlling hazardous activities boils down to a question of efficient ex ante *safety regulation. In addition, it is suggested that it is better to place AI in the existing legal categories and not to create a new electronic legal personality.*

I. Introduction

Artificial intelligence (AI) and recent breakthroughs in machine–human interactions and machine learning technology are affecting ever more aspects of our lives. AI technologies are not limited to increased pervasiveness, but are also characterised by continuous and surprising breakthroughs fostered by computation capabilities, algorithm design and communication technology.[1] AI is exponentially growing, and certain of its materialisations bring greater threats to privacy, are ethically questionable and are

* University of Ljubljana School of Economics and Business, Ljubljana, Slovenia; email: mitja.kovac@ef.uni-lj.si.
The author would like to thank Roger van den Bergh, Gerrit De Geest, Ben Depoorter, Matthew Dyson, Michael Faure, Paula Giliker, Paul Heald, Eric Helland, Jonathan Klick, Anne Lafarre, Alain Marciano, Philip Morgan, Jens Prüfer, Giovanni Ramello, Wolf-Georg Ringe, Hans-Bernd Schäfer, Ann-Sophie Vandenberghe, Bruce Wardhaugh, the participants of the IMA Workshop, University of York, 2020, the workshop session at the European Master in Law and Economics (EMLE) Midterm Meeting, Hamburg, 2019, the 110th Society of Legal Scholars Annual Conference at the University of Central Lancashire, Preston, 2019 and the participants of the AGCOM workshop on "Law and economics of big data and artificial intelligence", Rome, 2018 for their thoughtful comments, suggestions and advice. Funding received from: Slovenian Research Agency (Javna Agencija za Raziskovalno dejavnost Republike Slovenije, ARRS); name of the research project: Challenges of inclusive sustainable development in the predominant paradigm of economic and business sciences, grant no.: P5-0128.

[1] S Russell and P Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Prentice Hall, NJ, Pearson 2016) pp 2–5.

European Journal of Risk Regulation, 13 (2022), pp. 94–113
© The Author(s), 2021. Published by Cambridge University Press
doi:10.1017/err.2021.28

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

possibly even dangerous, risky and could cause potential catastrophic risk.[2] Whether to pursue the creation of non-natural AI[3] able to make choices via an evaluative process is one of the most pressing questions in the world today. Namely, AI[4] is unleashing a new industrial revolution, and it is vital that lawmakers systemically address its challenges and regulate its economic and social effects while not stifling innovation.

Russell, for example, argues that no one can predict exactly how the new AI technology will develop but, if autonomous machines start to far exceed our thinking capacity and we leave this issue unaddressed, AI could well be the last event in human history.[5] He suggests that poorly designed autonomous machines might pose a serious risk to humanity.[6] Moreover, Turner argues that AI is creating a growing legal vacuum in almost every domain touched by this unprecedented technological "development".[7] Similarly, Buyers suggests that lawyers are currently flummoxed as to what should happen when a self-driving car has a software failure and hits a pedestrian, or a drone's camera happens to catch someone skinny-dipping in a pool or taking a shower, or a robot kills a human in self-defence.[8] Furthermore, Teubner shows that AI agents may actually pose three new liability risks: (1) autonomy risk, stemming from standalone "decisions" taken by the AI agents; (2) association risk, arising from the close cooperation between people and AI agents; and (3) network risk, which occurs when computer systems are closely integrated with other computer systems.[9] In addition, AI might lead to serious indirect or direct harm.[10] For example, high-speed trading algorithms that could destabilise the stock market or cognitive

[2] RJ Sawyer, "Robot ethics" (2007), 318(5853) Science 1037; and P Lin, K Abney and GA Bekey, *Robot Ethics: The Ethical and Social Implications of Robotics* (Cambridge, MA, MIT Press 2011).

[3] D Weinbaum and V Veitas, "Open ended intelligence: the individuation of intelligent agents" (2017) 29(2) Journal of Experimental and Theoretical Artificial Intelligence 371–96.

[4] In this paper, the term "artificial intelligence" denotes autonomous AI that is independent and has the capacity to self-learn, interact, take autonomous decisions, develop emergent properties and adapt its behaviour/actions to the environment and has no life in a biological sense. In other words, AI's "behaviour" is determined by computer code that allows some room for "decision-making" by the machine itself, and the AI's behaviour is not entirely under the control of human actors. See, eg, Russell and Norvig, supra, note 1, 23–28; D McAllester and D Rosenblitt, "Systematic nonlinear planning" (1991) 2 AAAI-91 634–39; EJ Horowitz, JS Breese and M Henrion, "Decision theory in expert systems and artificial intelligence" (1988) 2 International Journal of Approximate Reasoning 247–302; and EJ Horswill, "Functional programming of behaviour-based systems" (2000) 9 Autonomous Robots 83–93. See also P Stone, R Brooks, E Brynjolfsson, R Calo, O Etzioni, G Hager et al, *Artificial Intelligence and Life in 2030* (Report of the 2015 study panel 50, Stanford University 2016); P McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (Chico, CA, AK Press 2004) p 133.

[5] S Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (London, Allen Lane 2019) p 4.

[6] ibid.

[7] J Turner, *Robot Rules: Regulating Artificial Intelligence* (London, Palgrave Macmillan 2019) pp 81–86.

[8] J Buyers, *Artificial Intelligence: The Practical Legal Issues* (Minehead, Law Brief Publishing 2018) pp 21–35.

[9] G Teubner, "Digital personhood? The status of autonomous software agents in private law" (2018) *Ancilla Iuris*. See also A Koch, "Liability for emerging digital technologies: an overview" (2020) 11(2) Journal of European Tort Law 115–36.

[10] See, eg, Russel and Nordig, supra, note 1; T Simonite, "AI software learns to make AI software" (2017) MIT Technology Review; Y Wilks, *Artificial Intelligence: Modern Magic or Dangerous Future?* (London, Icon Books 2019); JN Kim, MI Jordan and S Sastry, "Autonomous helicopter flight via reinforcement learning" (2004) Advances in Neural Information Processing Systems 16 (NIPS 2003); and M Minsky, *The Emotion Machine: Common-Sense Thinking, Artificial Intelligence, and the Future of the Human Mind* (New York, Simon & Schuster 2006).

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

radio systems that might interfere in emergency communications may, either alone or in combination, cause serious damage.[11]

Given such developments, many lawmakers around the globe have started on intensive law-making activity[12] with respect to the issue of liability and other broader challenges brought by emerging digital technologies. For example, the European Commission established a special expert group on liability made up of the "Product Liability Directive" formation and the "New Technologies" formation.[13] These two expert groups have also been given the task of determining whether regulatory intervention on AI technologies is appropriate and necessary and, if so, whether such an intervention should be developed in a horizontal or sectoral way.[14] Moreover, the issues of legal liability for AI and related civil liability for damages caused by AI have produced an impressive amount of scholarly literature, turning them into subjects of major interest for lawyers.[15]

By incorporating the main insights from the tort law and economics literature,[16] this paper joins this critical debate and offers an additional set of arguments on the appropriate role of the civil liability regime in regulating AI. This paper complements my earlier work on judgment-proof robots in two noteworthy respects.[17] First, the paper contributes to the literature by highlighting the practical and theoretical importance of the AI-related judgment-proof problem. Second, this paper focuses on the judgment-proofness of existing legal persons engaged with AI, who might not be held responsible for substantial harm as the current liability-related tort law stipulates.

The analysis presented here is both positive and normative. The analytical approach employs interdisciplinary analysis enriched with concepts used in the economic analysis

[11] Callvano et al show that the algorithms consistently learn to charge supra-competitive prices without communicating with each other; E Calvano, G Calzolari, V Denicolo and S Pastorello, "Algorithmic pricing: what implications for competition policy? (2019) 55(2) Review of Industrial Organization 155–71; and E Calvano, G Calzolari, V Denicolò and S Pastorello, "Artificial intelligence, algorithmic pricing, and collusion" (2020) 110(10) American Economic Review 3267–97. See also JE Harrington, "Developing competition law for collusion by autonomous artificial agents" (2018) 14(3) Journal of Competition Law & Economics 331–63.

[12] See, eg, Resolution on the Civil Law Rules on Robotics of the European Parliament, P8-TA (2017)0051.

[13] EU Commission, COM (2018) 237 final.

[14] ibid.

[15] For a synthesis, see S Lohsse, R Schulze and D Staudenmayer, "Liability for artificial intelligence", in S Lohsse, R Schulze and D Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Baden-Baden, Nomos Verlagsgesellschaft 2018) p 11; J De Bruyne and C Vanleenhove, *Artificial Intelligence and the Law* (Cambridge, Intersentia 2021); A Koch, "Liability for emerging digital technologies: an overview" (2020), 11(2) Journal of European Tort Law 115–36.

[16] See, eg, A Galasso and H Luo, "punishing robots: issues in the economics of tort liability and innovation in artificial intelligence" in *NBER Chapters, The Economics of Artificial Intelligence: An Agenda* (Cambridge, MA, National Bureau of Economic Research 2018) pp 493–504; HB Schäfer and C Ott, *The Economic Analysis of Civil Law* (Cheltenham, Edward Elgar 2004) pp 107–273; HB Schäfer, "Tort law: general" in B Bouckaert and G De Geest (eds), *Encyclopedia of Law and Economics* (Cheltenham, Edward Elgar 2000) pp 569–96; W Emons and J Sobel, "On the effectiveness of liability rules when agents are not identical" (1991) 58(2) Review of Economic Studies 375–90; S Shavell, *Economic Analysis of Accident Law* (Cambridge, MA, Harvard University Press 1987); AM Polinsky and WP Rogerson, "Product liability, consumer misperceptions and market power" (1983) 14(1) Bell Journal of Economics 581–89; S Shavell, "Strict liability versus negligence" (1980) 9(1) Journal of Legal Studies 1–25; RA Posner, "A theory of negligence" (1972) 1(1) Journal of Legal Studies 29–96; G Calabresi, "Some thoughts on risk distribution and the law of torts" (1961) 70(4) Yale Law Journal 499–553; G Calabresi, *The Costs of Accidents: A Legal and Economic Analysis* (New Haven, CT, Yale University Press 1970).

[17] M Kovac, *Judgement-Proof Robots and Artificial Intelligence. A Comparative Law and Economics Approach* (London, Palgrave Macmillan 2020).

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

of law.[18] However, several caveats should be issued. Namely, the limited scope of the paper considers the narrow fields of tort and product liability law while omitting analysis of consumer protection and anti-trust law. It solely focuses on European Union (EU) product liability and on AI that interacts with its environment in unforeseeable ways. Moreover, the aim of the paper is not to impose a final word on the matter, but to undertake an exploratory analysis of the relationship between AI and the judgment-proof problem.

This paper is structured as follows. The next section presents the general background and several definitions and provides a manual for the field of AI's development and deployment. In Section III, key questions for AI policy are considered. Moreover, crucial questions imposed by lawmakers are described, along with three critical fields of application. In Section IV, the paper provides several recommendations for lawmakers. Finally, some conclusions are presented.

II. General background and key concepts

Dartmouth College and the two-month workshop at Dartmouth in the summer of 1956 was to become the official birthplace of the AI field, and the following seventy years have seen a revolution in both the content and the methodology of work in AI.[19] Scientific progress has also enabled the return of neural networks[20] and the re-emergence of intelligent agents.[21] This re-emergence of intelligent agents suggests that previously isolated subfields of AI are to be tied together and has drawn AI into much closer contact with other fields, such as control theory and economics.[22]

1. Setting the scene: concepts and research trends

Influential founding fathers believed that AI should put less emphasis on creating applications that are good at performing specific tasks and should instead strive for machines that think, that learn and that create.[23] Closely related is the idea of artificial general intelligence (AGI) that looks for a universal algorithm for learning and acting

[18] See R Van der Bergh, *The Roundabouts of European Law and Economics* (The Hague, Eleven International Publishing 2018) pp 21–28; and RA Posner, *Economic Analysis of Law* (9th edn, Alphen aan den Rijn, Wolters Kluwer Law Publishers 2014).

[19] See eg J McCarthy, "From here to human-level AI" (2007) 171(18) Artificial Intelligence 1174–82; GF Luger, *Computation and Intelligence: Collected Readings* (Palo Alto, CA, AAAI Press 1995); J McCarthy, ML Minsky, N Rochester and CE Shannon, *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (Hanover, NH, Dartmouth College, tech. rep. 1955); and NJ Nillson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge, Cambridge University Press 2009).

[20] See, eg, P Smolensky, "On the proper treatment of connectionism" (1988) 11(1) Behavioral and Brain Sciences 1–74.

[21] Russel and Norvig, supra, note 1, at 26.

[22] ibid, at 27.

[23] See, eg, McCarthy, supra, note 19; ML Minsky, P Sigh and A Sloman, "Designing architectures for human-level intelligence" (2004) 25(2) AI Magazine 113–254; ML Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of Human Mind* (New York, Simon & Schuster 2007); N Nilsson, "Human-level artificial intelligence?" (2005) 26(4) AI Magazine 68–75; J Beal and PH Winston, "The new frontier of human-level artificial intelligence" (2009) 24(4) IEEE Intelligent Systems 21–23; and NJ Nilsson, *Artificial Intelligence: A New Synthesis* (Burlington, MA, Morgan Kaufmann 1998).

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

in any environment.[24] It is also helpful at the outset to introduce a distinction between narrow and general AI. Narrow AI, which also represents the focus of this paper, denotes the ability of a system to achieve a certain stipulated goal or set of goals, and the great majority of AI systems today are of this narrow type.[25] However, general AI denotes the ability to achieve an unlimited range of goals and even to set new goals independently.[26]

Furthermore, one must note the distinction between data-based and semantic AI systems. Semantic AI employs formal semantics to derive meaning from disparate sets of raw data into content.[27] This enables a computer system to have human-like understanding and reasoning. This is done by using tools, methods and techniques that help categorise and process data as well as define the relationships between different concepts and datasets. Thus, semantic technologies allow computers not only to process strings of characters, but also to store, manage and retrieve information based on meaning and logic.[28] The data-based AI systems are provided with a very large corpus of data (combined with learning methods), which enables the AI to learn new patterns, resulting in excellent performance.[29] In addition, it has to be emphasised that in recent years the field of AI has seen a shift from simply building systems that are intelligent to building intelligent systems that are human-aware and trustworthy.[30]

2. Literature review

The issue of an appropriate civil liability regime for AI has already produced an impressive amount of legal scholarship.[31] Some scholars investigate algorithms as

[24] See, eg, B Goertzel and C Pennachin, *Artificial General Intelligence* (Berlin, Springer 2007); E Yudkowsky, "Artificial intelligence as a positive and negative factor in global risk," in N Bostrom and M Cirkovic (eds), *Global Catastrophic Risk* (Oxford, Oxford University Press 2008); and S Omohundro, "The basis AI drives" (2008) AGI-08 Workshop on the Sociocultural, Ethical and Futurological Implications of Artificial Intelligence.

[25] See Weinbaum and Veitas, supra, note 3. However, see also M Boden, *AI: Its Nature and Future* (Oxford, Oxford University Press 2016) 119; and W Wallach and C Allen, *Moral Machines: Teaching Robots Right from Wrong* (Oxford, Oxford University Press 2009) p 68.

[26] Weinbaum and Veitas, supra, note 3.

[27] It combines the advantages of semantic reasoning and neural networks; M Acosta, P Cudré-Mauroux, M Maleshkova, T Pellegrini, H Sack and Y Sure-Vetter (eds), *Semantic Systems. The Power of AI and Knowledge Graphs* (Berlin, Springer 2019).

[28] T Poonam, TV Prasad and M Singh, "Comparative study of three declarative knowledge representation techniques" (2010) 2(7) International Journal of Advanced Trends in Computer Science and Engineering 2274. See also Nillson, supra, note 19.

[29] Data-based AI actually solves the "knowledge bottleneck" in AI (the problem of how to express all of the knowledge that a system needs): A Halevy, P Norvig and F Pereira, "The unreasonable effectiveness of data" (2009) 24(2) IEEE Intelligent Systems 8–12; R Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (New York, Viking Press 2005); and M Banko and E Bril, "Scaling to very very large corpora for natural language disambiguation" (2001) ACL-01: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, 26–33.

[30] See J Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford, Oxford University Press 2016); and P McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (Chico, CA, AK Press 2004) p 133.

[31] See, eg, R Leenes and F Lucivero, "Laws on robots, laws by robots, laws in robots: regulating robot behaviour by design" (2014) 6 Law, Innovation and Technology 193; U Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts* (Berlin, Springer 2013); PM Asaro, "A body to kick, but still no soul to damn: legal perspectives on robotics", in P Lin (ed.), *Robot Ethics: The Ethical and Social Implications of Robotics* (Cambridge, MA, MIT Press 2012) p 169; FP Hubbard, "'Sophisticated robots': balancing liability, regulation, and innovation", (2015) 66 Florida Law Review 1803; R de Bruin, "Autonomous intelligent cars on the European intersection of liability and privacy" (2016) 7(3) European Journal Risk Regulation 485–501; MF Lohmann, "Liability issues concerning self-driving vehicles" (2016) 7(2) European

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

such,[32] whereas others explore multiple set of issues and jurisdictions.[33] Amongst legal scholars there also appears to be a general preference for some form of strict liability for algorithms and robots, analogous to liability for animals or movable objects (or the liability for motorised vehicles).[34] Providing a comprehensive comparative analysis, Tjong Tjin Tai suggests that there are three areas that may require change: strict liability, product liability for algorithms and extending the protected interests.[35] Wagner argues that the main focus of current liability rules and the legal practice developed under them is on the users of technical appliances, not on the manufacturers.[36] Yet Wagner also suggests that the manufacturer who determines the safety features and the behaviour of the robot or Internet of Things device "clearly is the cheapest cost avoider, in fact, he is the only person in a position to take precautions at all".[37]

On the other hand, Abbott and Sarch discuss the difficulties involved in punishing AI and offer modest expansions to criminal law, including, most importantly, new negligence crimes centred around the improper design, operation and testing of AI applications, as well as possible criminal penalties for designated parties who fail to discharge statutory duties.[38] Rachum-Twaig suggests that current law and doctrine, such as product liability and negligence, cannot provide an adequate framework for these technological advancements, mainly due to the lack of personhood and agency and to the inability to predict and explain robot behaviour.[39] He argues that the inherent lack of foreseeability is challenging basic principles in tort law, which requires foresight prior to imposing liability.[40] Moreover, product liability doctrine "seems to struggle with the lack of foreseeability characterizing AI-based robots,

Journal Risk Regulation 335–40; and S Lohsse, R Schulze and D Staudenmayer, "Liability for artificial intelligence", in S Lohsse, R Schulze and D Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Baden-Baden, Nomos 2018) p 11.

[32] S Chopra and LF White, *A Legal Theory for Autonomous Artificial Agents* (Ann Arbor, MI, University of Michigan Press 2011); T Schulz, *Verantwortlichkeit bei autonom agierenden Systemen* (Baden-Baden, Nomos 2014); and EAR Dahiyat, "Towards new recognition of liability in the digital world: should we be more creative?" (2011) 19(3) International Journal of Law and Information Technology 224–42.

[33] See, eg, E Palmerini and A Bertolini, "Liability and risk management in robotics," in R Schulze and D Staudenmayer (eds), *Digital Revolution: Challenges for Contract Law in Practice* (Baden-Baden, Nomos 2016) p 225; and E Tjong Tjin Tai, "Aansprakelijkheid voor robots en algoritmes" (2017) Nederlands Tijdschrift voor Handelsrecht 123.

[34] See, eg, T Schulz, *Verantwortlichkeit bei autonom agierenden Systemen* (Baden-Baden, Nomos 2014); J Hanisch, *Haftung für Automation* (Göttingen, Cuvillier 2010); and S Gless and K Seelmann (eds), *Intelligente Agenten und das Recht* (Baden-Baden, Nomos 2016).

[35] Tjong Tjin Tai argues that strict liability for robots (and possibly algorithms) would have to be adopted (via specific statue) and imposed on the owner and/or user. He also suggests that product liability could be extended to algorithms (via statute); E Tjong Tjin Tai, "Liability for (semi)autonomous systems: robots and algorithms" in V Mak, E Tjong Tjin Tai and A Berlee (eds), *Research Handbook in Data Science and Law* (Cheltenham, Edward Elgar 2018) pp 55–82.

[36] G Wagner, "Robot liability" in S Lohsse, R Schulze and D Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things: Munster Colloquia on EU Law and the Digital Economy IV* (Baden-Baden, Nomos 2019).

[37] Thus, as Wagner suggests, "in the interest of meaningful incentives of the manufacturer to employ available safety measures and to balance their costs and benefits, manufacturer liability is essential"; ibid.

[38] R Abbott and A Sarch, "Punishing artificial intelligence: legal fiction or science fiction" (2019) 53(1) UC David Law Review 323–84.

[39] O Rachum-Twaig, "Whose robot is it anyway?: liability for artificial-intelligence-based robots" (2020) 2020(4) University of Illinois Law Review 1141–76.

[40] ibid.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

preventing a swift application of the design defect doctrine".[41] While exploring the deep normative structures of our societies, Eidenmüller argues that it would dehumanise the world if we were to treat machines like humans, even though machines may be smart – possibly even much smarter than humans.[42]

Moreover, Borghetti shows that broad liability regimes that have been designed to handle damage caused by humans are ill-suited for the compensation of harm caused by, or associated with, the use of AI.[43] However, Borghetti advances that sector-specific liability regimes or compensation mechanisms are applicable and do not require that an abnormal behaviour or conduct be established.[44] Wendehorst, for example, argues that AI-driven robots in public spaces should be subject to strict liability for damage resulting from their operation and that AI manufacturers should be liable for damage caused by defects in their products, even if the defect was caused by changes made to the product under the producer's control after it had been placed on the market.[45] However, Cabral suggests that the current EU Product Liability Directive is not up to the task of regulating AI, and it can neither adequately protect consumers nor foster innovation.[46]

III. A SYNTHESIS OF LAW AND ECONOMICS SCHOLARSHIP: TORTS AND SAFETY REGULATION

This section presents a set of law and economics recommendations that might shed light on the improved deterrence of hazards and the inducement of optimal precautions while simultaneously keeping dynamic efficiency – incentives to innovate – undistorted.

1. The economic function of tort law

Tort law defines the conditions in which a person is entitled to compensation for damage if not based on a contractual obligation and encompasses all legal norms that concern the claim made by an injured party against a wrongdoer (tortfeasor). Economically speaking, any "reduction of an individual's utility level caused by a tortious act can be regarded as

[41] ibid. However, De Bruyne and Vanleenhove argue that in relation to AI the existing rules of jurisdiction and applicable law do not pose particular problems when applied to self-driving cars; J De Bruyne and C Vanleenhove, "The rise of self-driving cars: is the private international law framework for non-contractual obligations posing a bump in the road?" (2018) 5(1) IALS Student Law Review 14–26.

[42] H Eidenmüller, "Machine performance and human failure: how shall we regulate autonomous machines?" (2019) 15(1) Journal of Business & Technology Law 109–33. See also E Karner, "Liability for robotics: current rules, challenges, and the need for innovative concepts," in S Lohsse, R Schulze and D Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Baden-Baden, Nomos 2018) p 117.

[43] Borghetti also argues that fault is not a relevant concept when algorithms are at stake, and establishing an algorithm's defect will probably be too difficult in most cases; JS Borghetti, "Civil liability for artificial intelligence: what should its basis be?" (2019) 17 Revue des Juristes de Sciences Po 94–102. See also JS Borghetti, "How can artificial intelligence be defective?" in S Lohsse, R Schulze and D Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Baden-Baden, Nomos 2018) p 63.

[44] Borghetti (2018). supra, note 43. See also P Machnikowski, "Producers' liability in the EC Expert Group report on liability for AI" (2020) 11(2) Journal of European Tort Law 137–49.

[45] C Wendehorst, "Strict liability for AI and other emerging technologies" (2020) 11(2) Journal of European Tort Law 150–80.

[46] TS Cabral, "Liability and artificial intelligence in the EU: assessing the adequacy of the current Product Liability Directive" (2020) 27(5) Maastricht Journal of European and Comparative Law 615–35.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

damage".[47] A thorough overview of the tort law and economics literature exceeds the limits of this paper and is available elsewhere.[48] However, it should be emphasised that this literature traditionally addresses three broad aspects of tortious liability. The first is assessing its incentives (including incentives to participate in activities and incentives to mitigate the associated risk) – analytically speaking, tort law is thus an instrument that improves the flow of inducements[49]; the second concerns risk-bearing capacity and insurance; while the third is related to the necessary administrative expense entailing the costs of legal services, the value of litigants' time and related lost opportunities and the court operating costs.[50] The literature also shows that since the administrative and procedural costs of a tort law case can be very high, alternative legal mechanisms such as *ex ante* safety regulation might be more cost-effective at reducing the overall costs of accidents.[51]

2. Liability for harm versus safety regulation

In his seminal paper on liability for harm versus the regulation of safety, Shavell paved the way to an analytical understanding of the optimal employment of tort liability and/or regulatory standards.[52] Shavell instrumentally addressed the effects of liability rules and direct regulation on a rational self-interested party's decision-making process.[53] Namely, liability in tort and safety regulation are two different approaches to controlling activities that create risks of harm and inducing the optimal amount of precaution.[54] Yet Shavell stressed that major mistakes have occurred in the use of liability and safety regulation.[55] Regulation, where applied exclusively, has for various reasons often proven inadequate, whereas due to causation problems tort liability might also provide suboptimal deterrence incentives.[56] In addition, Rose-Ackerman suggests that regulation (statutes) should generally dominate provided agencies are able to employ rule-making to shape policy,[57] whereas Schmitz argues that the joint use of liability and safety regulation is optimal if wealth varies among injurers.[58]

[47] G Calabresi and AD Melamed, "Property rules, liability rules and inalienability: one view of the cathedral" (1972) 85(6) Harvard Law Review 1089–128.

[48] See, eg, R Cooter and T Ulen, *Law and Economics* (6th edn, Boston, MA, Addison-Wesley 2016) pp 287–373; Posner, supra, note 16, at 14; HB Schäfer and C Ott, *The Economic Analysis of Civil Law* (Cheltenham, Edward Elgar 2004) pp 107–273; and E Mackaay, *Law and Economics for Civil Law Systems* (Cheltenham, Edward Elgar 2015).

[49] G De Geest, "Who should be immune from tort liability?" (2012) 41(2) Journal of Legal Studies 291–319.

[50] S Shavell, "Liability for accidents" in MA Polinsky and S Shavell (eds), *Handbook of Law and Economics* (Vol. 1, Amsterdam, North Holland 2007) pp 139–83.

[51] DN Dewees, D Duff and MJ Trebilcock, *Exploring the Domain of Accident Law: Taking the Facts Seriously* (Oxford, Oxford University Press 1996) p 452.

[52] For an overview of his contributions, see S Shavell, *Economic Analysis of Accident Law* (Cambridge, MA, Harvard University Press 2007).

[53] S Shavell, "Liability for harm versus regulation of safety" (1984) 13(2) Journal of Legal Studies 357–74.

[54] ibid.

[55] ibid.

[56] ibid. Also see RA Epstein, "The principles of environmental protection; the case of Superfund" (1982) 2(1) Cato Journal 9–53.

[57] S Rose-Ackerman, "Tort law as a regulatory system" (1991) 81 AEA Papers and Proceedings 2.

[58] PW Schmitz, "On the joint use of liability and safety regulation" (2000) 20(3) International Review of Law and Economics 371–82.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

3. Liability issues and the classic human-centric judgment-proof problem

In its original, narrow meaning of the concept human-centric "judgment-proof", the problem refers to the fact that human tortfeasors are unable to pay fully for the harm they may cause, giving them a bigger incentive than otherwise to engage in risky activities. Shavell and Summers coined the term "judgment-proof" in their path-breaking articles on this problem where they showed that the judgment-proof problem's very existence seriously undermines the deterrence and insurance goals of tort law. Shavell notes that judgment-proof parties do not have the right incentive to either prevent accidents or purchase liability insurance.[59] In other words, the judgment-proof problem is critical because if injurers are unable to pay in full for the harm they have caused, then their incentives to participate in risky activities will be greater than otherwise. Summers also shows that judgment-proof injurers tend to take too little precaution under strict liability since accident costs are only partly internalised.[60]

Moreover, one should note that strict liability provides incentives for the optimal engagement in an activity if "parties' assets are enough to cover the harm they might cause, but their incentives will be inadequate if they are unable to pay for the harm".[61] Furthermore, Shavell argues that, also under the negligence rule in situations where injurers are not induced to take optimal care (or there are errors in the determination of negligence), the "existence of the judgment-proof problem induces injurers to engage more frequently (sub-optimally) in the activity than they normally would".[62]

In addition, when injurers are for any reason unwilling to pay for all of the harm caused by virtue of complex asset ownership-shifting arrangements put in place in advance of a risky activity, this fact alone also distorts their incentive to make optimal precautionary and damage mitigation decisions, including distorting or even completely eliminating any reason to purchase liability insurance.[63] Namely, risk-averse injurers who may not be able to pay for all of the harm they cause will tend not to purchase full liability insurance or any at all.[64] Here, Shavell notes that "the nature and consequences of this judgement-proof's effect depend on whether liability insurers have information about the risk and hence link premiums to that risk".[65] Consequently, "reduction in the

[59] J Summers, "The case of the disappearing defendant: an economic analysis" (1983) 132 University of Pennsylvania Law Review 145–85; and S Shavell, "The judgement proof problem" (1986) 6(1) International Review of Law and Economics 45–58.

[60] Summers, supra, note 59.

[61] Shavell, supra, note 59. Also see JJ Ganuza and F Gomez, "Being soft on tort. Optimal negligence rule under limited liability" (2005) UPF Working paper; and J Boyd and DE Ingberman, "Noncompensatory damages and potential insolvency" (1994) 23(2) Journal of Legal Studies 895–910.

[62] Shavell, supra, note 50, at 148.

[63] Shavell offers an example of the injurer's problem of choosing care x under strict liability, when their assets are $y < h$ and where the injurer's problem is formulated as minimising $x + p(x)y$; where the injurer chooses $x(y)$ determined by $-p'(x)y = 1$ instead of $-p'(x)h = 1$, so that $x(y) < x^*$ (and the lower is y, the lower is $x(y)$). In this instance, the injurer's wealth after spending on care would be $y - x$, and only this amount would be left to be paid in a judgment; Shavell, supra, note 50, at 148.

[64] G Huberman, D Mayers and C Smith, "Optimal insurance policy indemnity schedules" (1983) 14(2) Bell Journal of Economics 415–26. Also see WR Keeton and E Kwerel, "Externalities in automobile insurance and the underinsured driver problem" (1984) 27(3) Journal of Law and Economics 149–79; and Shavell, supra, note 59.

[65] Shavell, supra, note 50, at 180.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

purchase of liability insurance tends to undesirably increase incentives to engage in the harmful activity".[66]

4. The judgment-proof problem in the context of artificial intelligence

The classic law and economics concept of the judgment-proof problem informs us that if injurers lack sufficient assets to pay for the damage they cause, then their incentives to reduce risk will be inadequate.[67] Yet, the judgment-proof problem could also be defined much more broadly to include the problem of the dilution of incentives to lower risk that emerge when an existing legal person engaged with AI is completely indifferent to both the *ex ante* possibility of being found legally liable for harm done to others and potential accident liability (given that the value of the expected sanction equals zero). In other words, existing legal persons might be completely indifferent to the *ex ante* possibility of being found liable by the human-imposed legal system for harm caused, and hence their incentives to engage in risky activities might be inadequate. For example, since the actions of AI agents are likely to become increasingly unforeseeable, designers or producers of AI might think that such unforeseeable development might excuse them from any tortious liability and, in such a scenario, the classic tort law mechanism might (except for at a very high level of abstraction and generality) become inadequate to deal with potential harm caused by AI agents.[68] It must be stressed that this problem of diluted incentives (a broad judgment-proof definition) is distinct from what scholars and practitioners often call a "judgment-proof problem", generally described as when a tortfeasor is merely financially unable to pay for all of the losses, leaving the victim without full compensation.[69] Thus, the judgment-proof characteristics of the existing legal persons engaged with AI might potentially undermine the deterrence and insurance goals of classic tort law. Namely, the evolution of AI and its capacity to develop characteristics in a manner never envisaged by its designers or producers could undermine the effectiveness of the traditional strict liability and other tort law instruments. The prospect that AI might behave in ways that designers or manufacturers did not expect challenges the prevailing assumption within tort law that courts only compensate for foreseeable injuries.

The judgment-proof characteristic also implies that AI's activity levels will tend to be socially excessive and will contribute to excessive risk-taking by the existing legal

[66] In addition, the problem of excessive engagement in risky activities is mitigated to the extent that liability insurance is purchased, but the problem of suboptimal levels of care could be exacerbated if the insurers' ability to monitor care is imperfect; see Shavell, supra, note 50, at 180.

[67] Shavell, supra, note 59, at 58.

[68] Moreover, such a liability might result in over-deterrence of such an AI data provider, the operator or a software engineer, and may be detrimental to innovation and also hamper innovation activity. See, eg, M Porter, *The Competitive Advantage of Nations* (New York, Free Press 1990); WP Viscusi and MJ Moore, "Product liability, research and development, and innovation" (1993) 101(1) Journal of Political Economy 161–84; J Pelkmans and A Renda, "Does EU regulation hinder or stimulate innovation?" (2014) Centre for European Policy Studies, Special report No. 26; and A Galasso and H Luo, "Risk-mitigating technologies: the case of radiation diagnostic devices" (2020) Management Science 1–19.

[69] See, eg, G Huberman et al, supra, note 64; and Keeton and Kwerel, supra, note 64.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

persons associated with AI.[70] They might have no liability-related incentive to mitigate risk and their incentives to reduce the risk and harm will be completely diluted. The deterrence goal might be corrupted irrespective of the liability rule since the judgment-proofness of existing legal persons associated with AI[71] might not *ex ante* internalise the costs of any accident such AI might cause. Hence, as the literature suggests, tortious liability might fail to provide adequate incentives to alleviate the risk.[72] In other words, the insurance goal will be undermined to the extent that the judgment-proof tortfeasor proves unable to fully compensate their victims. Moreover, as shown by Logue, first-party insurance markets also will not provide an adequate remedy.[73]

It has to be emphasised that AI is applied in many different settings, implying that different liability systems will also have to be applied for each sector. However, for most technological ecosystems there is no specific liability regime.[74] This means that product liability, general tort law rules (fault-based liability, tort of negligence, breach of statutory duty) and possibly contractual liability occupy centre stage.[75] Generally speaking, the potential independent development and self-learning capacity of an AI agent might thus cause its de facto immunity from tort law's deterrence capacity and consequential externalisation of the precaution costs.

5. Legal problems, artificial intelligence and liability causing judgment-proof problems

The existing liability frameworks that could conceivably apply to AI-generated consequences can be broken down (apart from contract law) into two distinct categories: tortious liability (negligence, strict liability) and product liability under consumer protection legislation. As to the former, current laws of tortious liability rely on concepts of causality and foreseeability. In common law systems,[76]

[70] See S Shavell, *Foundations of Economic Analysis of Law* (Cambridge, MA, Harvard University Press 2004) pp 175–289; R Pitchford, "Judgement-proofness" in P Newman (ed.), *The New Palgrave Dictionary of Economics and the Law* (London, Palgrave Macmillan 1998) pp 380–83; and AH Ringleb and SN Wiggins, "Liability and large-scale, long-term, hazards" (1990) 98(3) Journal of Political Economy 574–95.

[71] See, eg, G Corfield, "Tesla death smash probe: neither driver nor autopilot saw the truck" (2017) *The Register*; and S Levin and JC Wong, "Self-driving Uber kills Arizona women in first fatal crash involving pedestrian" (2018) *The Guardian*.

[72] Evidently, autonomous systems are expected to decrease the number and severity of accidents dramatically, but accidents will continue to occur. The critical point is that the pool of accidents that an autonomous system still causes will not be the same as the pool of accidents a reasonable driver is unable to avoid. However, as Wagner points out, "AI might fail to observe and account for a freak event that any human would have recognized and adapted his or her behaviour to"; Wagner, supra, note 36.

[73] KD Logue, "Solving the judgement-proof problem" (1994) 72 Texas Law Review 1375–94.

[74] Expert Group on Liability and New Technologies and New Technologies Formation, "Liability for Artificial Intelligence and Other Emerging Technologies", European Union, 2019.

[75] ibid.

[76] One has to note that civil law countries and the EU Member States operate their own liability systems, and the differences among these systems are manifold. Yet generally where an actor fails to take due care and this negligence causes harm to another or where a wrongdoer causes such harm intentionally, this actor is liable to compensate the victim. "The principle of fault-based liability covers harm done to a set of fundamental interests of the person, i.e. life, health, bodily integrity, freedom of movement, and private property; in some legal systems the list of protected interests also includes purely economic interests and human dignity"; Wagner, supra, note 36. For thorough analyses, see C von Bar, *The Common European Law of Torts* (Vol. 1, Munich, C.H. Beck 1998); and Koch, supra, note 15.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

foreseeability is, as Turner suggests, employed in establishing both the range of the potential claimants (was it foreseeable that this person would be harmed?) and the recoverable harm (what type of damage was foreseeable?).[77] Alfonseca et al report that the new AI uses purely unsupervised deep reinforcement learning, which does not require the provision of correct input/output pairs or any correction of suboptimal choices and is motivated by the maximisation of some notion of reward in an online fashion.[78] In principle, these representations may be difficult for humans to understand and scrutinise.[79] AI is becoming multifaceted and therefore potentially capable of mobilising a diversity of resources in order to achieve objectives that are potentially incomprehensible to humans, let alone controllable or foreseeable.[80]

Moreover, as Rahwan et al show, the ability of AI to "adapt using sophisticated machine learning algorithms makes it even more difficult to make assumptions about the eventual behavior of an AI".[81] Thus, the actions of AI are likely to become increasingly unforeseeable (ie the insolvability of the program-prediction problem),[82] and this could, as Karnow argues, challenge the prevailing assumption within common tort law that courts only compensate for foreseeable injuries.[83] Namely, AI may generate solutions that even an objective reasonable human being, no matter how optimal or experienced an observer they are, would not expect or may not have even considered.[84] Martin-Casals then suggests that if a particular legal system chooses to view the "experiences of some learning AI systems as so unforeseeable that it would be unfair to hold the system's designers liable for harm that these systems cause, victims might be left with no way of obtaining compensation for their losses".[85] However, one has to note that in such cases multiple tort law regimes apply simultaneously and that theoretical difficulties in establishing foreseeability are often in practice solved by flexible factual interpretations or evidentiary techniques.[86]

---

[77] Turner, supra, note 7. See also Expert Group on Liability and New Technologies and New Technologies Formation, supra, note 74, at 22–27; and M Infantino and E Zervogianni, "The European ways to causation," in M Infantino and E Zervogianni (eds), *Causation in European Tort Law* (Cambridge, Cambridge University Press 2017) pp 604–05.

[78] M Alfonseca, M Cebrian, AF Anta, L Coviello, A Abeliuk and I Rahwan, "Superintelligence cannot be contained: lessons from computability theory" (2021) 70 Journal of Artificial Intelligence Research 65–76. See also V Mnih, K Kavukcuoglu, SD Rusu, AA Veness, J Bellemare, MG Graves et al, "Human-level control through deep reinforcement learning" (2015) 518(7540) Nature 529–33.

[79] Alfonseca et al, supra, note 78.

[80] ibid. See also N Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford, Oxford University Press 2014).

[81] I Rahwan, M Cebrian, N Obradovich, J Bongard, J-F Bonnefon, C Breazeal et al, "Machine behaviour" (2019) 568(7753) Nature 477–86.

[82] While AI is the product of human creation, today the production process is so complicated that the producer or creator may be unable to predict the way in which the algorithm may respond to all possible input conditions; see Tjong Tjin Tai, supra, note 35.

[83] EAC Karnow, "The application of traditional tort theory to embodied machine intelligence," in R Calo, M Froomkin and I Kerr (eds), *Robot Law* (Cheltenham, Edward Elgar 2015).

[84] MU Schere, "Regulating artificial intelligence systems: risks, challenges, competencies, and strategies" (2016) 29(2) Harvard Journal of Law & Technology 353, at 363.

[85] M Martin-Casals, "Causation and scope of liability in the Internet of Things," in S Lohsse, R Schulze and D Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Baden-Baden, Nomos 2018) p 223.

[86] See, eg, Infantino and Zervogianni, supra, note 77, at 606; and H Kötz and G Wagner, *Deliktsrecht* (Berlin, Franz Vahlen 2016) p 94.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

On the other hand, legal systems are not entirely devoid of statutes governing extra-contractual liability. The two most developed systems of product liability are the EU's Product Liability Directive of 1985 (Council Directive 85/374/EEC)[87] and the US Restatement (Third) of Torts on Products Liability, 1997.[88] According to the EU Product Liability Directive, a product is defective "when it does not provide the safety which a person is entitled to expect, taking all circumstances into account, including (a) the presentation of a product; (b) the use to which it could reasonably be expected that the product would be put; (c) the time when the product was put into circulation".[89]

Thus, the literature suggests that the current EU Product Liability Directive might also suffer from the same shortcomings as the classic tort law system.[90] Namely, as suggested by the Expert Group on Liability and New Technologies and New Technologies Formation, the current Directive focuses on the moment when the product was put into circulation as the key turning point for the producer's liability, and this cuts off claims for anything the producer may subsequently add via some update or upgrade.[91] In addition, the EU Product Liability Directive does not provide for any duties to monitor the products after putting them into circulation.[92] Moreover, most EU Member States adopted the so-called development risk defence, which allows the producer to avoid liability if the state of scientific and technical knowledge at the time when they put the product into circulation was not such as to enable the existence of the defect to be discovered.[93] Furthermore, product liability regimes operate on the assumption that the product does not continue to change in an unpredictable manner once it has left the production line and, as shown, autonomous AI does not follow this paradigm.[94] However, one has to note that these potential legal challenges do not fundamentally hinder the application of the current Directive to AI producers. Namely, the current high bar for the development risk defence may actually exclude the unforeseeability of AI-related damages as a potential liability exception. Moreover, one may argue that such damages might not be regarded as unforeseeable since societies already know that AI has an autonomous potential that may cause all kinds of hazards.

[87] For thorough analyses, see Wagner, supra, note 36; Koch, supra, note 15; De Bruyne and Vanleenhove, supra, note 15; Palmerini and Bertolini, supra, note 33; Borghetti (2018), supra, note 43; and Lohsse et al, supra, note 15.

[88] M Shifton, "The Restatement (Third) of Torts: Products Liability – the Alps cure for prescription drug design liability" (2001) 29(6) Fordham Urban Law Journal 2343–86.

[89] For a thorough discussion on whether a piece of software is a product, see Wagner, supra, note 36.

[90] See Wagner, supra, note 36. See also Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the Application of the Council Directive on the approximation of the laws, regulations, and administrative provisions of the Member States concerning liability for defective products (85/374/EEC), COM(2018) 246 final, 8 f.

[91] Expert Group on Liability and New Technologies and New Technologies Formation, supra, note 74.

[92] ibid. However, such duties are contained in the general safety regulation and sector-specific legislation that is relevant within an AI context (see, eg, Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices). See also Wagner, supra, note 36; and Koch, supra, note 15.

[93] ibid.

[94] See Turner, supra, note 7, at 98. See also L Griffiths, P de Val and RJ Dormer, "Developments in English product liability law: a comparison with the American system" (1988) 62(354) Tulane Law Review 383–85.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

IV. Towards the optimal regulatory artificial intelligence intervention: what can law and economics offer lawmakers?

The previous discussion and application of the main findings of the law and economics literature to AI suggests that lawmakers might be facing the unprecedented challenge of simultaneously regulating possibly harmful and hazardous activity while not deterring innovation in the AI sector and associated industries. Yet, economically speaking, law is a much more resilient and robust mechanism than is often believed.[95] However, one may question whether the existing strict liability regimes are adequate to deal with the diluted incentives to reduce the risk associated with AI.[96] Thus, the classic debate on the two different means of controlling risks, namely *ex post* liability for harm done or *ex ante* safety regulation, may, due to the shortcomings of human-centred, liability-related tort law instruments, boil down to a question of efficient *ex ante* regulation.[97]

1. Policy suggestions to ameliorate the judgment-proof artificial intelligence problem

The law and economics literatures offer several potential types of policy responses to mitigate the identified judgment-proof problem. The first instrument is vicarious liability.[98] Shavell, for example, suggests that if another party (principal) has some control over the behaviour of the party whose assets are limited (agent), the principal can be held vicariously liable for the losses caused by the agent.[99] Hence, vicarious liability (indirect reduction of risk) and a specific principal–agent relationship between the owner (a human who uses AI) and their AI agent features as a satisfactory remedy for the AI-related risks. The principal (owner) should be held vicariously liable for the losses that the agent causes. If the principal can observe the agent's level of care, the imposition of vicarious liability will induce the principal to compel the agent to exercise optimal care. In other words, an extension of liability should indirectly lead to a reduction of risk.

How, then, would such a vicarious liability be applied to an AI agent? Turner offers an example of a police force that employs patrol AI agents that might, according to such a rule, be vicariously liable in instances where such a patrolling AI agent assaults an

[95] Namely, since all new technology in essence presents a certain conceptual problem to the existing jurisprudence, efficient legal institutions react and generally address such issues by, for example, requiring legal standards of reasonableness, duty of care or good faith. See, eg, Guille v. Swan, Supreme Court of New York 1822; and Rylands v. Fletcher (1868) LR 3 HL 330.

[96] See OJ Erdelyi and J Goldsmith, "Regulating artificial intelligence: proposal for a global solution" (2018) AIES 95–101; and V Wadhwa, "Laws and ethics can't keep pace with technology" (2014) 15 Massachusetts Institute of Technology: Technology Review.

[97] P Schmitz, supra, note 58. See also A Agrawal, J Gans and A Goldfarb, "Prediction, judgment, and complexity: a theory of decision-making and artificial intelligence," in A Agrawal, J Gans and A Goldfarb (eds), *The Economics of Artificial Intelligence: An Agenda* (Chicago, IL, University of Chicago Press 2019).

[98] For syntheses, see A Sykes, "The economics of vicarious liability" (1984) 93 Yale Law Journal 168–206; and RH Kraakman, "Vicarious and corporate civil liability," in G De Geest and B Bouckaert (eds), *Encyclopedia of Law and Economics* (Vol. II, Civil Law and Economics, Cheltenham, Edward Elgar 2000).

[99] Shavell, supra, note 59. Also see Shavell, supra, note 52.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

innocent person during its patrol.[100] Moreover, the unilateral or autonomous actions of AI agents that are not foreseeable do not necessarily operate (as in the instance of product liability) so as to break the chain of causation between the person held liable and the harm.[101] Yet the literature also shows that if the principal is unable to observe and control the level of care exercised by the agent (AI), then they will generally be unable to compel the agent.[102] Nevertheless, if the principal can control the AI's level of activity (but has no observation capacity), then such vicarious liability will induce the principal to reduce the AI's participation in risky activity.

However, what if AI is truly autonomous, being capable of self-learning, developing emergent properties and adapting its behaviour to the environment? In these circumstances, the imposition of vicarious liability might prove inadequate due to the extreme judgment-proof problem. Lawmakers should then combine strict liability and vicarious liability – the strict liability of the manufacturer and the vicarious liability of the principal (any existing legal person). Moreover, the identified judgment-proof problem also implies that the current option proposed by the EU Expert Group on Liability for New Technologies and New Technologies Formation to introduce vicarious liability for autonomous systems[103] in order to address the risks of emerging digital technologies might fall short of attempted deterrence and prevention goals. In order to mitigate the identified shortcomings of vicarious liability, the literature[104] offers the following instruments.

First, lawmakers could require any principal to have a certain minimum amount of assets in order to be allowed to engage in an AI-related activity.[105] Such an amount of assets then acts as an insurance for the acts of AI agents, induces principals to take precautions and serves as a mechanism to indirectly mitigate the judgment-proof problem.[106] Pitchford, for example, suggests that partial lender liability and an equivalent minimum equity requirement deliver the highest level of efficiency.[107] However, as Shavell points out, such a minimum asset requirements may also undesirably prevent some individuals who ought to engage in AI-related activity from doing so.[108]

---

100 Turner, supra, note 7.

101 ibid.

102 Shavell, supra, note 50, at 180.

103 Expert Group on Liability and New Technologies and New Technologies Formation, supra, note 74, at 45–46.

104 See, eg, Shavell, supra, note 70; and Shavell, supra, note 50, at 139–83.

105 Similar to the required minimum starting capital for corporations.

106 Shavell shows that principals will engage in the activity if and only if their benefits would exceed the expected harm caused; and if they engage in the activity, they will choose the optimal level of care. If individuals' assets are less than the potential harm, however, they will engage too often in the harmful activity, as they will not then face (effective) expected liability equal to the expected harm, and they will similarly lack incentives to take optimal care; see Shavell, supra, note 50.

107 R Pitchford, "How liable should a lender be? The case of judgement-proof firms and environmental risk" (1995) 85 American Economic Review 1171–86.

108 Namely, although their assets are low and their care would be inadequate, their benefits might still exceed the expected harm that they create; see Shavell, supra, note 50, at 170. Shavell also suggest that minimum asset requirements are somewhat blunt instruments for alleviating the incentive problems; see S Shavell, "Minimum asset requirements and compulsory liability insurance as solutions to the judgement-proof problem" (2005) 36(1) Rand Journal of Economics 63–77.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

Second, lawmakers could introduce the compulsory purchase of liability insurance coverage in order for any principal to be allowed to engage in autonomous AI-related activity.[109] Such insurance coverage would provide *ex ante* incentives for optimal precaution and for optimal principals' decisions as to whether to engage with superhuman AI-related activity at all.[110] For example, AI developers or users seeking coverage for an agent could submit it to a certification procedure and, if successful, would be quoted with an insurance rate depending on the probable risks posed by the AI agent.[111] However, one has to note that liability insurance requirements tend to improve parties' incentives to reduce risk when insurers can observe levels of care, but they dilute incentives to reduce risk when insurers cannot observe levels of care.[112] In the former case, if principals/users purchase full liability insurance coverage, their incentives to reduce risk would be optimal; in the latter case, compulsory liability insurance may be inferior to minimum asset requirements.[113] Moreover, if insurers indeed cannot observe AI-related risk and the moral hazard exists, then mandating the purchase of liability insurance may not be desirable. In such circumstances, Shavell suggests that an opposite form of insurance regulation may be advantageous: barring the purchase of liability insurance.[114]

Third, lawmakers could directly *ex ante* regulate the AI's risk-creating behaviour. Namely, regulatory agencies could *ex ante* set detailed standards for the behaviour, employment, operation and functioning of any AI.[115] For example, the idea would be

[109] In fact, the European Parliament has already advised the European Commission to consider and adopt a mandatory insurance scheme with respect to robotics and AI; European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

[110] Potential injurers may make superior decisions as to whether to engage in an activity and, if they do so, may have stronger incentives to reduce risk when they have at stake at least the required level of assets and/or liability insurance coverage if they are sued for causing harm; ibid. Moreover, a party with assets less than the possible harm can pay at most their assets and thus faces a commensurately low expected liability. But, as Shavell suggests, if the party must purchase liability insurance in order to engage in the activity, they will bear a higher expected liability, and this may improve their decisions as to whether to participate in the activity; ibid. See also PJ Jost, "Limited liability and the requirement to purchase insurance" (1996) 16 International Review of Law and Economics 259–76.

[111] Karnow suggest that risk would be assessed along a spectrum of automation: the higher the intelligence, the higher the risk, and thus the higher the premium, and vice versa. If third parties declined to deal with uncertified programs, the system would become self-fulfilling and self-policing. Sites should be sufficiently concerned to wish to deal only with certified agents. Programmers (or others with an interest in using, licensing or selling the agent) would in effect be required to secure a Turing certification, pay the premium and thereby secure protection for sites at which AI agents are employed; CEA Karnow, "Liability for distributed artificial intelligences" (1996) 11(147) Berkeley Technology Law Journal 193–94. Interestingly, such a system was already put forth back in the days of slavery to account for the autonomous acts of slaves – admittedly a discomforting comparison; JB Wahl, "Legal constraints on slave masters: the problem of social cost" (1997) 41(1) American Journal of Legal History 1–24.

[112] Shavell, supra, note 108.

[113] ibid.

[114] Forbidding the purchase of liability insurance can then improve incentives to take care if, without a prohibition, AI users or developers would have purchased positive coverage and insurers cannot observe the injurer's level of care; ibid. See also MK Polborn, "Mandatory insurance and the judgement-proof problem" (1998) 18(2) International Review of Law and Economics 141–46.

[115] Shavell points out that such direct regulation – safety standards – will help to form incentives for the principals and the manufacturer to *ex ante* reduce risk as a precondition for engaging in an activity; see Shavell, supra, note 108. See also BW Smith, "Automated driving and product liability" (2017) 2017(1) Michigan State Law Review 1–74; and KS Abraham and RL Rabin, "Automated vehicles and manufacturer responsibility for accidents: a new legal regime for a new era" (2019) 105(1) Virginia Law Review 127–71.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

to simply regulatorily limit the AI's abilities in order to prevent it from doing harm to humans.[116] Such *ex ante* regulations and safety pre-emptions would also significantly reduce the degree of uncertainty regarding liability risk, and this, in general, increases research and development.[117] Furthermore, harmonising different, slow-moving Member States-wide regulations could also speed up experimentation and safe AI adoption.[118]

Fourth, regulatory agencies could set a detailed set of sector-specific safety standards[119] (similar to those in the air travel or pharmaceutical industries).[120] For example, under the existing rules, AI must meet essential health and safety requirements,[121] and efforts to produce harmonised European standards for AI are ongoing.[122] Such standards could, for example, require a special driver's license to operate a self-driving car.[123] Similarly, doctors may be required to take a minimum number of training sessions with a robotic system before being allowed to perform certain types of procedures on patients.[124] The literature shows that such safety standards (and related liability for the breach of these safety standards) may incentivise users themselves to innovate in ways that help them to take more effective precautions and may incentivise producer innovation because users would demand safer and easier-to-use design features, and mandatory training would favour "easier-to-teach" designs in order to reduce adoption costs.[125]

However, one has to note that such safety standards alone are inadequate and should be combined with the *ex ante* registration of both the principal and the superhuman AI agents (ie Turing registries). Such all-encompassing registries, like that for vehicles or ships, decrease information asymmetries, enable more effective regulatory control of hazardous activities and act as efficient *ex ante* mechanisms to deter and prevent disastrous events.[126]

---

[116] Yet such an intervention might simply forgo the enormous potential benefits of AI. See J Babcock, J Kram'ar and RV Yampolskiy, *The AI Containment Problem* (Berlin, Springer 2016) pp 53–63.

[117] Kaplow suggests that the basic trade-offs depend on factors including the frequency and the degree of heterogeneity of adverse events, as well as the relative costs of individuals in learning and applying the law; L Kaplow, "Rules versus standards: an economic analysis" (1992) 42 Duke Law Journal 557–629.

[118] Galasso and Luo, supra, note 16.

[119] Scherer suggests establishing a regulatory authority dedicated to regulating and governing the development of AI; see Scherer, supra, note 84.

[120] This also implies the establishment of a specialised superhuman AI regulator, an agency encompassing all superhuman AI-related activities (similar to the US Food and Drugs Administration (FDA)).

[121] See, eg, Directive EC 2006/42 on machinery; Directive 2014/53/EU on radio equipment; Directive 2001/95/EC on general product safety.

[122] EU Commission, COM (2018) 237 final.

[123] Such regulation would actually maintain AI use–consumer liability to the extent that users of AI technologies have sufficient incentives to take precautions and invest in training, thus internalising potential harm to others; see Galasso and Luo, supra, note 16, at 499. See also B Hay and K Spier, "Manufacturer liability for harms caused by consumers to Others" (2005) 95 American Economic Review 1700–11.

[124] Galasso and Luo, supra, note 16, at 499. See also B O'Reilly, "Patents running out: time to take stock of robotic surgery" (2014) 25 International Urogynecology Journal 711–13.

[125] See E Von Hippel, *Democratizing Innovation* (Cambridge, MA, MIT Press 2005); Hay and Spier, supra, note 123; and Galasso and Luo, supra, note 16.

[126] Yet it has to be emphasised that such regulation may involve inefficiency because of regulators' limited knowledge of risk and of the cost and ability to reduce it; see Shavell, supra, note 50, at 171.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

Fifth, criminal liability for the principal[127] could be introduced in order to provide additional pressure to optimise the principal's decision as to whether to engage with the AI activity at all. Namely, a principal who would not take care if only their assets were at stake might be induced to do so for fear of imprisonment.[128]

Sixth, lawmakers could extend liability from the actual injurer (the AI) to the company that engages or employs such an AI agent. Such an extension of liability could be achieved by piercing the veil of incorporation, for example.

Seventh, lawmakers could introduce corrective *ex ante* taxes that would equal the expected harm. Such corrective taxes, while implying the *ex ante* internalisation of potential damages (negative externalities), would then *ex ante* induce the optimal level of activity and AI-related engagement. Namely, when harm is caused with a low probability, the expected harm is much less than the actual harm, and parties with limited assets may be able to pay the appropriate tax on risk-creating behaviour even though they could not pay for the harm itself.[129] For example, owners, developers or users of AI – or just certain types of AI – could pay a tax into a fund to ensure adequate compensation for victims of AI crime.[130]

Eighth, lawmakers could establish a regime of compulsory compensation[131] or a wide insurance fund for instances of catastrophic losses that is publicly and privately financed. Such insurance implies a risk-sharing and risk-pooling mechanism (throughout the entire society) and is the optimal risk allocation in instances of unforeseeable, unpreventable catastrophic harms.[132] One should note that such insurance schemes already exist in the nuclear industry.[133] For example, the Price Anderson Act for nuclear power establishes a pool of funds to compensate victims in the event of a nuclear incident through a chain of indemnity regardless of who was ultimately at fault.[134]

---

[127] Abbot and Sarch argue that existing criminal law coverage will in cases of hard AI crimes likely fall short, and that additional AI-related offenses must be created to adequately deter novel crimes implemented with the use of AI; Abbot and Sarch, supra, note 38. Moreover, Hallevy explains that it "seems legally suitable for situations in which an AI entity committed an offense, while the programmer or user had no knowledge of it, had not intended it, and had not participated in it"; see G Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems* (Berlin, Springer 2015).

[128] See, eg, People v Davis 958 P.2d 1083 (Cal. 1998). For a restatement of such a liability with regards to joint enterprise criminal liability in the UK, see R v. Jogee, Ruddock v. The Queen (2016) UKSC8, (2016) UKPC 7.

[129] Shavell illustrates this principle with an example of pollution release that will cause harm of $1 million with a 1% probability. Thus, a firm with $100,000 of assets would be able to pay $10,000 for the expected harm it would cause were it to cause the pollution, even though it would only be able to pay one-tenth of the actual $1 million harm it might generate, and so its incentives to reduce risk would be much too low under the liability system; see Shavell, supra, note 50, at 171.

[130] Abbot and Sarch, supra, note 38.

[131] Giufrida et al suggest that this could be modelled, for instance, on the International Oil Pollution Compensation Funds, created under the auspices of the International Maritime Organization pursuant to the 1992 International Convention on Civil Liability for Oil Pollution Damage and the 1992 International Convention on the Establishment of an International Fund for Compensation for Oil Pollution Damage; see I Giuffrida, F Lederer and N Vermeys, "A legal perspective on the trials and tribulations of AI: how artificial intelligence, the Internet of Things, smart contracts, and other technologies will affect the law" (2018) 68(3) Case Western Reserve Law Review 747–81.

[132] An AI compensation fund could, for example, operate like the National Vaccine Injury Compensation Program (VICP). Namely, vaccines create widespread social benefits but are known in rare cases to cause serious medical problems. VICP is a no-fault alternative to traditional tort liability that compensates individuals injured by a VICP-covered vaccine. It is funded by a tax on vaccines that is paid by users. See National Vaccine Injury Compensation Program <https://www.hrsa.gov/vaccine-compensation/index.html> (last accessed 8 February 2021).

[133] Moreover, New Zealand has replaced tort law with a publicly funded insurance scheme to compensate victims of accidents. See, eg, PH Schuck, "Tort reform, Kiwi-style" (2008) 27(1) Yale Law & Policy Review 187–90.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

In addition, one should consider introducing the AI manufacturer's strict liability supplemented by the requirement that an unexcused violation of a statutory safety standard is negligence per se. Moreover, compliance with the regulation standard should not relieve the injurer's principal from tort liability. Thus, the rule per se (violation of a regulatory standard implies tort liability – including strict liability) should also be applied to AI-related torts, and the compliance defence of an AI manufacturer or its principal should not be accepted as an excuse.[135] One has to note that regulation and tort law should be applied simultaneously. *Ex post* liability and *ex ante* regulation (safety standards) are generally viewed as substitutes for correcting externalities where the usual recommendation is to employ the policy that produces lower administrative costs. However, Schmitz shows that the joint use of liability and regulation can enhance social wealth.[136] Namely, regulation removes problems that affect liability while liability limits the cost of regulation.[137] That is, by introducing an *ex ante* regulatory standard, the principal might be prevented from taking low levels of precaution and might find it convenient to comply with the regulatory standard despite the judgment-proof problem.[138]

2. A new special electronic legal person should not be created

Regarding the specific legal status in paragraph 59 of its Resolution on Civil Law Rules in Robotics,[139] the European Parliament suggests that the EU create a specific legal status for robots so that at least the most sophisticated autonomous robots can be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying an electronic personality to cases where robots (AI) make autonomous decisions or otherwise interact with third parties independently. Moreover, Solum,[140] Wright,[141] Teubner[142] and Koops et al[143] also argue that AI should be given a legal personality and that there is no compelling reason to restrict the attribution of action exclusively to humans and social systems. Furthermore, Allen and Widdison state that when an AI is capable of developing its own strategy it makes sense that the AI should be held responsible for its independent actions.[144] Yet it must be emphasised that Teubner,

[134] The Price–Anderson Act, background info (Center for Nuclear Science & Technology Information, La Grange Park, IL), November 2005. See also Abbot and Sarch, supra, note 38.

[135] G De Geest and G Dari-Mattiacci, "Soft regulators, tough judges" (2007) 15(2) Supreme Court Economic Review 119–40.

[136] Schmitz, supra, note 58.

[137] S Rose-Ackerman, "Regulation and the law of torts" (1991) 81 American Economic Review 54–58.

[138] De Geest and Dari-Mattiacci, supra, note 135.

[139] P8_TA (2017) 0051.

[140] LB Solum, "Legal personhood for artificial intelligences" (1992) 70 North Carolina Law Review 1231.

[141] GR Wright, "The pale cast of thought: on the legal status of sophisticated androids" (2001) 25 Legal Studies Forum 297, at 297.

[142] G Teubner, "Rights of non-humans? Electronic agents and animals as new actors in politics and law" (2007) Lecture delivered on 17 January 2007, Max Weber Lecture Series MWP 2007/04. Also see Teubner, supra, note 9.

[143] BJ Koops, M Hildebrandt and DO Jaquet-Chiffell, "Bridging the accountability gap: rights for new entities in the information society?" (2010) 11(2) Minnesota Journal of Law, Science & Technology 497–561.

[144] T Allen and R Widdison, "Can computers make contracts?" (1996) 9(1) Harvard Journal of Law & Technology 25–52.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press

for example, suggests that software agents should be given a carefully calibrated legal status.[145] The solution to the risk brought by autonomy would, according to Teubner, be their status as actants, as actors with partial legal personhood whose autonomous decisions are made legally binding in case they trigger liability for damages.[146]

Obviously, from a law and economics perspective, the establishment of a special status of an electronic person for AI that would have its own legal personality and responsibility for potential damages should be avoided. Namely, the establishment of such a legal personality would actually amplify the existing judgment-proof problem. It would completely dilute the incentives of the existing legal persons engaged with AI to reduce the risk that arises due to their complete indifference both to the *ex ante* possibility of being found legally liable for harm (as the liability now falls upon the AI) done to others and to the potential accident liability (where the value of the expected sanction equals zero). In other words, the establishment of such a specific electronic person might institutionalise the judgment-proofness of existing legal persons engaged with AI. Accordingly, the establishment of an unregulated human-like electronic personality is not an effective or adequate response to the identified AI-related judgment-proofness, but might be seen as an amplifier making the problem even more persistent. Consequently, granting legal personality to autonomous AI might open Pandora's Box of moral hazard and create perverse incentives on the side of human principals, the AI industry, designers, users and owners, and it would exacerbate the AI judgment-proof problem.

V. Conclusions

This paper sought to address the role of public regulatory policy in regulating AI and the related risk and civil liability for damage caused by such AI. As argued, existing legal persons associated with AI in their daily enterprises might be completely indifferent to the *ex ante* possibility of being found liable by the human-imposed legal system for harm caused, and hence their incentives to engage in risky activities might be socially excessive. The judgment-proof characteristic also implies that it might induce excessive risk-taking by the existing legal persons associated with AI. Thus, the identified judgment-proof characteristics may undermine the deterrence and insurance goals of tort law. In order to mitigate the identified effects of the judgement-proofness of AI-associated companies, a specific set of *ex ante* regulatory interventions is suggested.

Furthermore, the contemplated new special electronic legal personality for AI should not be introduced. As this paper attempts to show, the judgment-proofness of AI implies that any establishment of a legal personality would, while exacerbating the judgment-proof problem, bring about unexpected adverse effects. This paper also shows that, due to the identified shortcomings, the debate regarding the different ways of controlling hazardous activities may be reduced to a question of efficient *ex ante* safety regulation. In other words, regulatory intervention is, from the law and economics perspective, the best option for governing AI systems.

[145] Teubner, supra, note 9. See also Expert Group on Liability and New Technologies and New Technologies Formation, supra, note 74.

[146] ibid.

https://doi.org/10.1017/err.2021.28 Published online by Cambridge University Press