EFFICIENT ACCOUNTABILITY REGIME FOR AI AND MACHINE LEARNING

Monika

EMLE Thesis 2020-21, Aix-Marseille University, Aix-en-Provence

Keywords: Artificial Intelligence, Regulation, Tort, Liability, Economic Efficiency

JEL Code: K290, O330

Authorship Declaration

I hereby declare and confirm that this thesis is entirely the result of my own work except otherwise indicated. I acknowledge the supervision and guidance I have received from Pierre Bentatta and Nathalle Rubio. This thesis is not used as part of any other examination and has not yet been published.

11.08.2021/Monika/ 

1. Abstract

Artificial Intelligence ("**AI**"), in simple terms, refers to technology of analysing large data sets and arriving at an automated decision. While, AI promises great future in terms of smart decision-making, achieving tasks which are humanly impossible and greater user convenience. At the same time, increasing use of AI has made people wary of its consequences as unpredictability of AI decisions poses legitimate legal issues. One of the issues is the accountability of parties involved in AI systems with respect to the harm arising from AI actions. This not only impacts AI consumer interest, but also general trust on AI. This thesis examines contract law and tort law economics to find efficient accountability regime for AI. It also sees how insurance impacts the liability of AI. Furthermore, it briefly looks at regulatory approach of selected public and private parties.

2. Acknowledgement

I would like to extend my sincerest gratitude to Prof. Pierre Bentatta for his insightful teachings and guidance. Also, Prof. Pierre Garello for being a constant guide.

Furthermore, all the resources provided by Aix-Marseille University, Erasmus University and Hamburg University for the completion of this thesis. I would also like to thank my fellow students who provided useful critic and suggestions.

Thank you.

3. Table of Content

1. Introduction

In March 2018, in Arizona a self-driving car was test-driving at night at the speed of 64 Kmph.[1] The car algorithm spotted "*unknown object"* on the road. It then identified it as another vehicle and then as a bicycle.[2] Meanwhile, the car hit the "*unknown object".* Later, it was discovered that the "*unknow object"* was a woman crossing the road.[3] It was found that the safety driver was also using her mobile phone and did not act in time to avoid the accident.[4] However, it was also argued that due to incorrect assessment by AI, control could not be given to the safety driver in time.[5]

This is an example of transformative nature of the Artificial Intelligence (AI) as well as the dangers that come with it. AI can drive cars, diagnose diseases, make paintings, trade shares, write books, play chess and much more.[6] If we observe, AI is already ubiquitous. It is described as a transformative technology with profound effect on "*manufacturing, robotics, transportation, agriculture, modeling and forecasting, education, cybersecurity, and many other applications."*[7] A 2018 McKinsey research states that by 2030, 70 percent of companies would have adopted at least one type of AI technology and it can potentially deliver around $13 trillion to global economy.[8] In

[1] Michael Li, *Another Self-Driving Car Accident, Another AI Development Lesson*, 16 November 2019, Available at: https://towardsdatascience.com/another-self-driving-car-accident-another-ai-development-lesson-b2ce3dbb4444 (Last accessed on 4 August 2021).
[2] Ibid.
[3] Ibid.
[4] Ibid.
[5] Ibid.
[6] Luke Dormehl, Thinking Machines: The Quest for Artificial Intelligence and Where It's Taking Us Next, Tarcher Perigee, 2017, p.8.
[7] John Villasenor, *Products liability law as a way to address AI harms*, Brookings, 31 October 2019, Available at: https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms/ (Last accessed on 4 August 2021).
[8] Jacques Bughin, et. al, *Notes from the AI frontier: Modelling the impact of AI on the world economy*, McKinsey Global Institute, 4 September 2018, Available at: https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-

fact, Stuart Russell, a leading Artificial Intelligence (AI) expert, described super intelligent AI as "*Success would be the biggest event in human history . . . and perhaps the last event in human history*."[9]

At the same time, it is unpredictable and poses new risks to human life, safety and well-being. Many are skeptical of the risks of applying AI.[10] These risks arise from the unique feature of AI which also makes it a promising technology. While risks of robots taking over humans may be overblown, it cannot be ignored that autonomous nature of AI reduces (in some cases, eliminates) human involvement. A natural corollary of this situation, is to question if we need to relook at our liability laws. Whether existing liability regime with respect to accidents and product harms would suffice? There have been some steps towards it[11], but legal uncertainty remains. Furthermore, Galasso and Luo argue that right balance of liability will be important for the diffusion of technology in the field of AI and robotics.[12] Another paper analyzed US car market and concluded that with strictness of liability regime, investment in AVs decreases.[13] Therefore, careful consideration of legal liability of parties involved in AI related accident or harm is needed.

---

frontier-modeling-the-impact-of-ai-on-the-world-economy (Last accessed on 4 August 2021).

[9] Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*, Viking, 2019, p.11.

[10] Benjamin Cheatham, Kia Javanmardian, and Hamid Samandari, *Confronting the risks of artificial intelligence*, 26 April, 2019, McKinsey Global Institute, Available at: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence (Last accessed on 4 August 2021).

[11] Jessica Fjeld et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Berkman Klein Center Research Publication No. 2020-1, January 15, 2020, Available at SSRN: https://ssrn.com/abstract=3518482 or http://dx.doi.org/10.2139/ssrn.3518482 (Last accessed on 4 August 2021).

[12] Alberto Galasso and Hong Luo,*When does product liability risk chill innovation? evidence from medical implants*, National Bureau of Economic Research, No. wp 25068, September 2018.

[13] Herbert Dawid and Gert Muehlheusser, *Smart Products: Liability, Timing of Market Introduction and Investments in Product Safety*, Bielefeld University, 2019.

1.1. Research Questions

In this context, it is necessary to understand how AI-human interactions can harm human beings and how it should be dealt with. In other words, how AI should be held accountable to minimize the risk of harm and at the same time benefit from its qualities. Therefore, thesis frames the following research questions:

a. What are the regulatory proposals to make AI accountable for the potential harms?
b. What is the efficient liability regime to address the harms posed by the AI?

Under **first research question,** this thesis conducts comparative analysis of public and private actors' approach to the AI accountability and liability. It identifies the issues of consensus and differences among these actors. For this purpose, thesis takes positive analytical approach and examines policy statements, declarations and regulatory proposals put forward by governments as well as private companies. **Second research question**, requires normative analysis on the basis of law & economics principles. In this regard, thesis looks at the risks posed by AI and challenges in addressing those risks. Then, it inspects various law & economics approaches suitable for containing AI risks and allow innovation, and attempts to ascertain most suitable approach.

1.2. Scope

There are many interlinked issues which stem from the usage of AI. However, this thesis will not address all of them. **First**, the distinction between AI as goods or service or a hybrid, is not relevant here. Wherever, thesis refers to AI or AI systems, it includes AI software, AI as service, any product or service based on AI. **Second**, there are various technical classification of AI, Machine Learning (ML), and Robots. Also, AI can be classified into various types on the basis of degree of autonomy. However, for

the purposes of this thesis AI or AI system refers to a machine system which has ability to learn and evolve autonomously and continuously with experience. Definition of AI will be dealt in detail in further sections. **Third**, debate if AI should be the subject or object of rights, is outside the scope of the thesis. While AI has ability of becoming a super-intelligent entity and can be argued that it can have legal rights and duties. However, for the purpose of the thesis, AI does not have a personhood and is treated as such. **Fourth**, one of the harms arising from AI is bias and discrimination. This thesis will not look into ethical and legal implications of that. It will focus on physical, psychological, and/or economic harms caused to users or third parties while using AI systems.

Finance, health, national security, governance, and consumer goods-all these sectors have consumers in some form and harm caused by AI is a concern for all of them. Nevertheless, one perspective, to understand the potential harm caused by AI and responses to that, is from consumer goods, including transportation. Autonomous vehicles can be the most commonly heard example of advanced applications of AI. This thesis will use that as an example to explain the relationship between AI developer, Product (Car) Manufacturer/Producer, Consumer, and Regulator. Learnings from this case would be applicable to all goods where consumers consume any good or service based on AI and there is risk of harm through usage. Accountability of AI in such a case is relevant for all consumer goods as well as for other applications like Finance, Health etc.

1.3. Structure of Thesis

The Section 2, explains characteristics of AI, risks posed by it and challenges to the current legal system. Next, Section 3, compiles and compares policy approach of various public and private actors towards AI accountability. Section 4, looks into tort

and contractual liability regime and its application to AI from the perspective of law and economics concepts. Finally, Section 5 draws the implications of policy approach discussed under Section 3 and present brief conclusion on efficient liability regime suited to AI.

2. Artificial Intelligence and its Challenges

2.1. Definition of Artificial Intelligence

John McCarthy was first one to use the term "artificial intelligence". He defined it as *"the science and engineering of making intelligent machines, especially intelligent computer programs".*[14] EU proposal on regulation for harmonized rules of artificial intelligence (EU Proposal) defines AI systems as "***software*** *that… can, for a* ***given set of human-defined objectives****, generate outputs such as content, predictions, recommendations, or decisions* ***influencing the environments*** *they interact with;"*[15] (emphasis supplied) Ebers views AI as a *"catch-all term"* which refers to the *"broad branch of computer science that studies and designs intelligent machines.*"[16] Russell and Norvig, categorize AI into four types: Thinking Humanly; Thinking rationally; Acting Humanly and Acting Rationally.[17] He goes on to state that there is no need for *"single all-encompassing definition"*, instead it is more important to understand the characteristics of AI and their practical application.*[18]*

---

[14] John McCarthy, *What is Artificial Intelligence?* 2007, Available at www.formal.stanford.edu/jmc/whatisai.pdf (Last accessed on 4 August 2021).
[15] European Commission, Proposal for a Regulation of the European Parliament and the Council Laying Down Harmonised Rules of Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 2021/0106, 21 April, 2021 (EU Proposal), Article 3.1.
[16] Martin Ebers, *Regulating AI and Robotics* in *Algorithms and Law*, Cambridge University Press, 2020, 37-99, p.41.
[17] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th edn., Pearson, 2021, p.1.
[18] Supra note 16, Ebers, p.42.

As per West and Allen, AI has intentionality, intelligence and adaptability.[19] Intentionality refers to the ability to collect data from various sources, analyze them in real time and act on the derived insights instantly.[20] As opposed to traditional machines which produce predetermined responses, AI can arrive at responses on the basis of new data collected. Intelligence refers to the ML. It allows AI to analyze variety of unstructured data sets and identify useful patterns.[21] Lastly, adaptability refers to the *"ability to learn and adapt as they make decisions".*[22] With the help of these features, AI supported self-driving car can collect data from various sources such as camera, sensors, Global Positioning System (GPS), then analyze the data in real time and take sophisticated decisions such as speed and direction of driving, avoiding clash with other objects on the road, etc.

From Ebers perspective, characteristics of AI which are significant for legal standpoint are: contractual complexity involving multiple parties; reliance of AI on correlation instead of causation; autonomy; and algorithms as black boxes.[23] These distinguishing characteristics will be discussed in more detail later in this section. Notably, this thesis uses AI as a "catch-all term" for all types of AI- strong or weak; AI or Artificial General Intelligence (AGI); ML; smart; robots; algorithmic systems; autonomous, etc. It stands for a computer system which has considerable degree of automation and self-learning capability.

[19] Darrell M. West and John R. Allen, *How artificial intelligence is transforming the world*, Brookings, 24 April 2018, Available at: https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/ (Last accessed on 4 August 2021).
[20] Ibid.
[21] Ibid.
[22] Ibid.
[23] Supra note 16, Ebers, pp.44-50

2.2. Risks of AI

However, as evident from the incident earlier, AI equipped "smart devices" are not always "smart". They are capable of misreading the situation and making faulty decision. These decisions can cause physical, economic and psychological harm to humans. Therefore, many are skeptical of the risks of applying AI.[24] Any AI is designed with a specific objective or predetermined goal. However, instead of successfully achieving that objective, AI may behave in unintended manner. It can either fail to achieve the objective or achieve the objective but in the process produce intended by-results. For example, a smart Roomba which is programmed to clean the house recognized 'dog feces' as something to be cleaned. But, in the process of cleaning the house it spread the feces all over the house.[25] In another case, IBM's AI program Watson was marketed to provide cancer treatment recommendations. However, it was reported to underperform and in fact, "cause sever and even fatal consequences".[26] Similarly, Amazon delivery drone can drop packet on the customer or any other person; smart house heating system can overheat or freeze and cause physical harm to the users; or AI based share trading can make bad decision and cause financial loss. Thus, with prevalence of AI based goods and services, occurrences of AI related harms is bound to increase too. These AI system or AI equipped product or service can cause harm to users or third parties which can be physical, psychological or economical in nature.

[24] Supra note 10, Cheatham.
[25] Olivia Solon, *Roomba Creator responds to reports of 'poopocalpse': We see this a lot*, 15 August 2016, Available at: https://www.theguardian.com/technology/2016/aug/15/roomba-robot-vacuum-poopocalypse-facebook-post (Last accessed on 4 August 2021).
[26] *2018 in Review: 10 AI Failures*, 11 December 2018, Available at: https://medium.com/syncedreview/2018-in-review-10-ai-failures-c18faadf5983 (Last accessed on 4 August 2021).

These harms or malfunctions can be caused by variety of factors. As per Amodei et al. a designer can specify a wrong objective; omit extrapolation for bad results; or rely on insufficient or poor-quality data inputs.[27] One of the key features of AI is "autonomy."[28] Thus, AI while bound by its predetermined objectives, can still behave in autonomous nature to achieve those objectives. Furthermore, its decision making is characterized by "blackbox" models. In case of ML, blackbox models are created by algorithms *"that humans, even those who design them, cannot understand how variables are being combined to make predictions."*.[29]

3. Current Regulatory Landscape

As AI is technologically progressing, discourse over principles of AI design and development has progressed. Consequently, several institutions- government, international organizations, civil society organizations and private enterprises, have put forward principles for governance of AI. These principles mostly touch upon privacy, fairness, human rights, safety & security etc. A report by Berkman Klein Center has identified eight thematic principles- Privacy; Accountability; Safety and Security; Transparency and Explainability; Fairness and Non-discrimination; Human Control of Technology; Professional Responsibility; Promotion of Human Values; and International Human Rights.[30] This report also maps global consensus on these themes.

[27] Dario Amodei, et al., *Concrete problems in AI safety*, 2016, Available at: *arXiv preprint arXiv:1606.06565* (Last accessed on 4 August 2021), pp.2-3
[28] European Commission, *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, 19 February 2020. (EU Safety Report)
[29] Cynthia Rudin and Joanna Radin, *Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition*, Harvard Data Science Review, 1(2) 2019.
[30] Supra note 11, Fjeld, pp.8-9.

It identifies multistakeholder; private sector; intergovernmental, governmental, and civil society organizations which have in some form declared all or some AI principles.[31]

This thesis, focuses on AI principles which have relevance for liability and accountability of AI in case of any harm originating from AI devices or services. It examines, in detail, how private and government sector has responded to AI accountability, respectively. For the purpose of this examination, we will look at 4 private and 4 government instruments or policies. These are- International Business Machines Corporation (IBM), Microsoft, Tencent, and Google from private sector, and Europe, United Kingdom, India and China from public sector.

[31] Ibid.

Table 1: AI Accountability Principles in Public and Private Policy

S. No.	Measure	Sector	Pre-Market Deployment						Post-Market Deployment and Pre-Incident				Post-Occurrence of Harm			
				Ex-Ante Measures									Ex-Post Measures			
			Prohibition	Registration	Certification	Technical Documentation	User Instruction	Conformity Assessment	Risk Management System/Audit	Log Keeping/Transparency	Human Control	Explainability	Negligence	Strict Liability	Joint & Several/Proportional Liability	Insurance
1.	Europe	Government	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✘	✘	✔	✘
2.	Unit	Gover	✔	✘	✘	✔	✘	✘	✘	✘	✘	✔	✘	✘	✘	✘

	ed King dom	nment														
3.	India	Gover nment	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✓	✗	✓	✗
4.	Chin a	Gover nment	✗	✗	✓	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗
5.	Tenc ent	Privat e	✓	✗	✗	✗	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗
6.	Micr osoft	Privat e	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗	✗	✓	✗

7.	Google	Private	✔	✘	✘	✔	✘	✔	✔	✔	✔	✔	✘	✘	✘	✘
8.	IBM	Private	✘	✘	✘	✔	✘	✘	✘	✔	✘	✔	✘	✘	✔	✘

Source: Author's Compilation

3.1. Regulatory Snapshot

Table 1 provides a snapshot of selected private and public actors approach to AI accountability.

General vs Specific References: Most actors mention the term "accountable AI", not all have mentioned the methods of implementation of that. Many actors use very general terms. Microsoft makes general reference to accountable, reliable & safe and transparent AI systems.[32] Similarly, IBM makes general reference to accountability and explainability.[33] However, it does specifically focus on boundaries of company or software responsibility vis-à-vis AI designer or developer. It puts onus on AI developer to know the company policy of accountability and other AI principles-fairness, user data rights etc. and reflect them in the AI system. Google as well presents very wide statement on safety and accountability of AI.[34] Tencent advocates for differential transparency due to intellectual property and technical nature of some AI systems.[35] It focuses more on explainability than technical transparency. It also promotes informational self-determination by providing users with sufficient information. It promotes effective control by humans, risk control measures and test and validation for reasonable expectation of performance. It adds that precautionary principle should be exercised and boundary of AI application should be defined. Thus, it agrees with exclusion of certain applications from AI. Tencent is only private player which has provide more specifications for AI accountability and safety.

---

[32] Microsoft, *AI Principles*, 2018, Available at: https://www.microsoft.com/en-us/ai/our-approach-to-ai (Last accessed on 4 August 2021).
[33] IBM, *IBM Everyday Ethics for AI*, 2019, Available at: https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf (Last accessed on 4 August 2021).
[34] Google, *AI at Google: Our Principles*, 2018, Available at: https://www.blog.google/technology/ai/ai-principles/ (Last accessed on 4 August 2021).
[35] Tencent Institute, *Six Principles of AI*, 2017, Available at: http://www.kejilie.com/iyiou/article/ZRZFn2.html (Last accessed on 4 August 2021).

Maturity of Regulation: Among, public sector players (national governments), EU has most detailed principles and a regulatory proposal. It is due to maturity of development of AI principles and governance framework. According to one study, AI governance is in its third stage since 2016: **first stage** involved publication of ethical principles by 85 technology companies and governments; **second stage** revolved around forming consensus on principles like: accountability, human control, explainability etc; **third and current stage** started in 2019. It is leading to conversion of AI ethical principles into adopted principles and policies.

The table shows parties at the third stage of AI governance. All these governments and technology company recognize accountability as an essential principle for AI development and deployment. While some parties are at nascent end of the third stage, others are at much advanced end. In this case, EU is most advanced with regulatory proposal and detailed enforcement of accountability through certification, registration, conformity assessment, etc. China is next. It has a detailed action plan for making China world leader in AI by 2030.[36]However, it focuses more on exploration of all AI technologies and investment for research and development, instead of AI governance. Thus, it mentions human-controlled technologies as one of the AI systems to be looked into. It does, however, recognize importance of testing, validation, certification and risk control. Next, United Kingdom (UK) is also at explorative stage of AI governance. It, while recognizing that technical transparency is not feasible in all cases, recommends

[36] Chinese National Governance Committee for the New Generation Artificial Intelligence, led by China's Ministry of Science and Technology, *Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence*, 2019, Available at: http://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html (Last accessed on 4 August 2021).

technical transparency in safety-critical scenarios and regulators. It also stresses on need to have "full and satisfactory explanations" for AI decisions.[37]

Ex-post Liability: India's AI principles are in form of a discussion paper.[38] It is only actor to mention ex-post liability instead of ex-ante measures. It proposes negligence or fault-based liability where all parties are proportionally liable. It also, mentions that safe-harbor may be created for certain AI applications. EU proposal instead focuses on ex-ante measures which are applicable pre and post deployment of AI. In case of violation of regulations, it imposes penalties. However, it has one similarity with Indian approach. It specifies duties and corresponding liability for all parties involved in AI use- high-risk AI systems providers (Article 16), product manufacturer (Article 24), AI systems importer (Article 26), distributor (Article 27), and user (Article 29).

There are some other notable observations:

Scope of Principles: EU proposal explicitly excludes certain applications of AI. The EU proposal "lays down…prohibitions of certain artificial intelligence practices;"[39]. These are AI systems that deploy *subliminal techniques beyond person's consciousness in order to distort the behavior*;[40] *exploit vulnerabilities of a specific group of persons*;[41] evaluate or classify *trustworthiness of natural persons* on the basis of their *social behaviour*;[42] resulting in physical and psychological harm, and use *real time biometric identification* for law enforcement.[43] Some other sectors such as civil aviation, vehicle

[37] Ibid, p.40.
[38] Niti Aayog, *National Strategy for Artificial Intelligence: #AI for All*, *2018,* Available at: https://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf (Last accessed on 4 August 2021).
[39] Supra note 13, EU Proposal, Article 1 (a).
[40] Supra note 13, EU Proposal, Article 5 (a).
[41] Supra note 13, EU Proposal, Article 5 (b).
[42] Supra note 13, EU Proposal, Article 5 (c).
[43] Supra note 13, EU Proposal, Article 5 (d).

and transportations and military applications are excluded.[44] UK, Tencent, and Google do not make such specific exclusion. Instead, it reiterates that AI applications should be beneficial to humanity. Tencent in fact provides for boundaries of AI applications.

Types of Harm: AI applications, as explained in examples in Section 2, can cause physical, psychological and economic harm. None of the actors in the table, except EU, mention type of harm that AI should prevent. EU also talk about "*preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system"*[45] It ignores economic-harm caused by AI applications. Furthermore, there can be harm to health or life by omission of AI. For instance, AI fails to diagnose a medical condition. Prima facie, such harms are covered as they can "emerge from AI". However, detection of such harms will be extremely difficult.

Differentiated obligations: Both EU and Tencent propose differentiated obligations with respect to AI. EU has more vigorous regulations for high-risk AI systems. Tencent on other hand proposes differentiated transparency requirement due to intellectual property and technical (e.g. blackbox) concerns.

Regulatory Structure: China proposes two-tier regulatory framework. EU framework can be considered three tier: It envisages European Artificial Intelligence Board at the helm (Article 56), national supervising authorities in Member states and then notified bodies to conduct market surveillance and testing etc.

4. Liability Regime for AI

The regulatory landscape discussed in the previous section mostly deals with the ex-ante measures. The goal of such measures is to prevent or minimize the risk of harm. Thus, through certification. safety standards, transparency, penalty for violation of

[44] Supra note 13, EU Proposal, Article 2.2 and 2.3.
[45] Supra note 13, EU Proposal, Article 14.2.

regulations, AI producers/providers/operators etc. are incentivized to avoid risky activities. However, such regulations are not always enough. Due to inability to detect violation or AI providers internalizing costs of penalty and deciding to profit from risky activity can render such regulations ineffective. Moreover, despite perfect implementation, accident or harm may still arise.

One scholar argues that for AI system harms, the legal responsibility for the operation of AI lies with the operator or a person who sets the parameters of the operation or the person who controls the behavior.[46] On the other hand, Hubbard claims, in the context of United States, that corrective justice system of tort and contract balance the liability and innovation, and no change is needed.[47] Similarly, Smith opines that the current product liability regime is compatible with automated driving systems.[48] Meanwhile, Liar concluded that strict liability is best suited to AI systems.[49]

Hence, issues of ex-post civil liability arise from use of AI systems. These also aim at creating disincentive for AI provider/operator etc. such that they prevent from risky activity and victim is compensated for the harm suffered. This can be done in two manners: contractual liability and tort liability.

4.1. Contract Liability

Now we will consider if contractual liability is sufficient to contain the risks of AI:

[46] Elena A. Kirillova, *Artificial Intelligence as a New Category of Civil Law,* Journal of Advanced Research in Law and Economics, 11(1), 2020.
[47] F. Patrick Hubbard, *Sophisticated Robots": Balancing Liability, Regulation, and Innovation,* Florida Law Review, 66(5), 2015.
[48] Bryant W. Smith, *Automated Driving and Product Liability,* Michigan State Law Review Journal, 1, 2017, 1-74.
[49] Anat Lior, *The AI Accident Network: Artificial Intelligence Liability Meets Network Theory,* Tulane Law Review 95, 2020.

The Coase theorem claims that "negative externalities" can be addressed efficiently by contracts, irrespective of assignment of rights, provided transaction cost is zero.[50]

Thus, contracts can be efficiently used to decide liability for the harms caused by AI system and therefore, internalize the social cost of AI system. For instance, for autonomous car, if harm is caused to the user (driver) due to mis-reading by the AI, then as per the user agreement between the user and AI provider, liability and compensation can be decided.

Multiple Parties: However, solution is not this straightforward. Any AI product involves multiple parties. For instance, AI developer, AI provider owning the AI, AI product manufacturer, AI operator after the deployment of the product, AI user, etc. While all these relations will have some form of contract among them, for AI user it becomes very complex.

When autonomous car gets into an accident, the user might have several agreements with different parties. There will be AI designer/developer; entity owning the AI; entity owning the car; entity operating the AI; entity collecting the data; etc. These entities can be same or different, depending on individual cases. In this circumstance, transaction cost for the user to discover who is the person responsible for the AI accident (Information Cost) is extreme. It has been observed that individual users are faced by large number of contractual partners- hardware, user agreements for software, digital content, software updates, end user license. [51] Thus, it creates complex web of contractual relationships. In addition, this exasperates the situation by making discovery of proof harm difficult.

---

[50] Ronald H. Coase, *The Problem of Social Cost*, Journal of Law & Economics 1(3) 1960.

[51] Supra note 16, Ebers, p.43; Christaine Wendehorst, *Sale of Goods in the Digital Age – From Bipolar to Multi-party Relationships* in The Age of Uniform Law. Essays in honour of Michael Joachim Bonell to celebrate his 70th birthday, UNIDROIT, 2016, 1873–1887.

Opacity and Autonomy: For AI, as mentioned earlier one of the distinguishing features is its autonomy and black box model. This opacity can also be due to concerns like: competitive advantage, [52] national security,[53]or privacy[54] or intellectual property rights. Hence, discovery of harm and its cause is not easy for the end user. This in addition to multiple contractual parties, can frustrate end user from discovering which party is at fault. [55] Since, burden of proof for breach of contract lies with person claiming the breach, complexity and opacity of AI based contracts can render any remedy futile. [56] In this situation there is an information asymmetry. In such scenario, ex-ante regulations for transparency and explainability can help discharge of burden of proof by the end user (victim).

Correlation over Causation: Furthermore, another challenge to contractual liability regime by AI is: correlation. as per Chris Anderson, AI works on correlation and not causation. He says, "Correlation of enough".[57] This is antithesis of concept of "causality" on which contractual liability is based.

Foreseeability: As regards foreseeability, we have already seen numerous instances of AI making decisions that a person would not have made or would have made differently.[58] For example, AlphaZero, an AI program owned by Alphabet, taught itself to play and defeat chess grandmaster. It discovered many unique approaches which

[52] Rob Kitchin, *Thinking Critically about and Researching Algorithms*, Information, Communication and Society 20(1) 2017, 1–14.
[53] Matthias Leese, *The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-discriminatory Safeguards in the European Union*, Security Dialogue, 45(5), 2014, 494-511.
[54] Brent D. Mittelstadt, et al., The Ethics of Algorithms: Mapping the Debate, Big Data & Society, December 2016.
[55] Supra note 16, Ebers, p.43; Supra note 51, Wenderhorst.
[56] Supra note 51, Wenderhorst.
[57] Chris Anderson, *The End of Theory,* Wired , July 2008.
[58] Supra note 16, Ebers, p.47.

were new to chess experts.[59] This problem is compounded by constant developing nature of AI systems. Since, AI systems learn at design phase and continue self-learning after deployment of the product in the market, it is impossible to foresee possible harms.[60]

These situations lead to high transaction costs of contract enforcement and in turn fails to internalize the cost of harm caused by the AI system. Thus, as per Coase Theorem, private bargaining is not the pareto efficient outcome in the case of AI. Also, in case of harm to third parties, again negotiation costs would be too high to make private bargaining inefficient for internalization of cost of harm.

However, in case of relationship between non-user parties involved in AI, contracts can still serve usefulness. For instance, in case of accident of an autonomous car, it leads to high transaction cost when contract is to be enforced between user and other AI parties. However, the transaction cost will be much lower for bargaining between AI developer or AI owner or AI owner- AI operator, etc. In fact, IBM stresses upon this demarcation of accountability between different AI parties.

Cheapest Cost Avoider: Furthermore, AI provider or AI owner which is most visible to the user, can be the one who is held liable. In case of Uber self-driving car, Uber is most visible party. Additionally, it is also arguably with maximum financial capacity. Therefore, it would be economically efficient, if user enforces the contract with respect to AI system harm with regard to Uber. Later, Uber can adjust its costs with other AI parties with whom it has contractual relationships. **First,** Uber is superior risk bearer and more immune to judgement-proof problem. **Second,** Uber is cheapest cost avoider. It is in better position to know and understand the contractual relationships with AI

[59] Will Knight, *Defeated Chess Champ Garry Kasparov Has Made Peace With A*I, 2 February, 2021, Available at: https://www.wired.com/story/defeated-chess-champ-garry-kasparov-made-peace-ai/ (Last accessed on 4 August 2021).
[60] Supra note 16, Ebers, p.47.

developer and operator. Thus, AI provider can have joint and several liability for the AI system contractual obligations towards a user.

Here, AI provider can pay the expectation damages to the user. As per bargain theory, *"promises should be enforced, according to the bargain theory, if they are part of a bargain, and the remedy for the breach of an enforceable promise is an award of the value expected of the bargain"*[61]

However, contractual bargaining is not sufficient for harm to third parties. If by stander is injured in self-driving car accident, then transaction costs are much higher to enter into a contract. Therefore, contract liability is not economically efficient remedy for redressal of AI system harms.

4.2. Tort Liability

The above-mentioned challenges of complexity, opacity, correlation, automation, foreseeability also equally applicable to tort liability.[62] In the case of AI systems we only focus on the unilateral liability cases. Since, in AI systems, user does not have much control over how AI system performs, user's negligence is insignificant. More control over AI performance is with AI developer, provider, or operator.

Furthermore, if harm is caused to a third party, then also we are looking at unilateral cases.

"*Economists describe harms that are outside private agreements as externalities. The economic purpose of tort liability is to induce injurers and victims to internalize the costs of harm that can occur from failing to take care."*[63] The tort law performs the internalization by inducing the injurer to compensate the victims.[64] Cooter and Ulen,

[61] Robert Cooter and Thomas Ulen, *Law and Economics,* 2016, p.281.
[62] Supra note 16, Ebers, p.43.
[63] Supra note 61, Cooter and Ulen, p.189.
[64] Supra note 61, Cooter and Ulen, p.190.

claim that "*the economic essence of tort law is its use of liability to internalize externalities created by high transaction costs*".[65]

The various liability standards governing AI systems are mentioned in the Table 2.

Table 2: Anatomy of Tort Liability Regime

Fault-based Liability without presumption/Negligence	**Fault-based liability with presumptions**	**Strict Liability**
Harm Causation	Rebuttable Presumption in favour of victim	Harm Causation Limiting principle
Injurer Friendly	Victim Friendly	Victim friendly
High Enforcement Cost	High Enforcement Cost	Lower Enforcement Cost Chilling Effect

In this table, three types of liability standards are mentioned. First, fault-based liability or negligence; second fault based-liability, but presumption in favor of the victim for burden of proof; third, strict liability or no-fault liability standard.

Negligence: The traditional theory of tort requires that the victim has suffered a harm, injurer has caused the harm by action or omission, and injurer had breached a duty owed to the victim.[66]

In AI systems, where the functioning of the system in veiled, victim is not in position to discover the harm unless it is very apparent. Especially, if harm is caused by omission. If AI system omits to give good financial advice or it omits detect a medical condition,

[65] Supra note 61, Cooter and Ulen, p.191.
[66] Supra note 61, Cooter and Ulen, p.191.

or smart Roomba omits to clean the house properly. In these examples, detection of harm by the user is negligible. Any action or omission emerging from AI action/omission, leads to harm if it reduces the utility function for the victim.[67]

Next, tort liability relies on establishing causality. It signifies that harm must be caused by the defendant/AI provider etc.[68] However, due to Blackbox model of the AI algorithm, victim who has suffered the harm does not have capability to discover the cause of harm. The traditional tests of but-for-test, or, left hand rule, etc. are irrelevant in the case of AI system.

Third component requires that defendant has a duty to care with regard to the AI and defendant breaches that duty. In other words, *"under a negligence rule, the defendant escapes liability if he satisfied the applicable standard of care to avoid the harm that he caused.*"[69] Like many cases, the EU proposal for regulation, imposes safety regulations on all AI players. These safety regulations are meant to be standard of care for determining if defendant performed as per negligent manner a lot. Thus, EU proposal can be used as a guide to negligence standard. However, despite that, negative externalities can still occur.

In addition, in case of fault-based liability, Scherer questions if AI autonomy can be considered "*an intervening force or act that is deemed sufficient to prevent liability for an actor whose tortious conduct was a factual cause of harm.*" [70] Furthermore, as automation increases, it becomes increasingly difficult to identify person who exercised

[67] Supra note 61, Cooter and Ulen, p.191.
[68] Supra note 61, Cooter and Ulen, p.193.
[69] Supra note 61, Cooter and Ulen, p.193.
[70] Matthew Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies and Strategies,* Harvard Journal of Law and Technology, 29(2), 2016, 354-400, pp. 363–364.

intention to cause harm, making it difficult to assign accountability and consequent liability.[71]

Thus, due to inherent nature of AI systems, negligence standard incurs high transaction costs. User as well as court do not have technical reach and know-how to decipher the blackbox of AI systems and determine the standard of care followed at each stage. Furthermore, if presumption in favor of the victim is introduced for causality or breach of duty, then the transaction costs are reduced. Nonetheless, issues related to AI systems still persist. Hence, under fault liability, cost of expected harm for the AI system provider is much lower than the cost of precaution.[72] Therefore, he/she has no incentive to exercise any precaution. Another consequence of this is that victim will be required to internalize the costs of AI harm. This will induce him/her to avoid risks and abstain from AI systems. This can have unintended repercussions of reduction of demand of AI systems. For instance, if self-driving cars keep getting into accidents and neither car manufacturer or AI operator are held liable, then along with physical and psychological harm, he /she will also bear the costs of damage caused to the car. This can force, many individuals to not to try the self-driving car. Such consequence has adverse effect on AI designing and development.

Strict Liability:

A rule of liability based upon harm and causation is called "strict liability."[73] The present EU system establishes strict liability (no-fault liability) for the producers of defective products in case they cause physical or material damage.[74] However, it covers tangible products, and not intangible products like AI. Also, many times it is difficult to

[71]Supra note 16, Ebers, p.56.
[72] Guido Calabresi, *The Costs of Accidents,* 1970.
[73] Supra note 61, Cooter and Ulen, p.196.
[74] Product Liability Directive 85/374/EEC, 1985.

classify AI as product or service. Additionally, under this Directive, injured party has to prove that product was defective at the time of circulation. Since, AI is self-learning and evolving in nature and continues to do so after the circulation, it conflicts with time of defect discovery/knowledge. In contrast, under a rule of strict liability, proof of causation is a necessary condition for liability, and proof of negligence is unnecessary.

Therefore, accordingly, transaction cost of successful prosecution comes down, social cost for the injurer rises, incentive to exercise greater precaution is instilled.

Thus, on the basis of this discussion and deficiencies of present contractual and product liability regulation, there is need to see what should be liability standard for AI systems.

4.3. What's Suitable?

From economic perspective, tort liability forces the injurer to internalize the cost of harm caused by his/her activities.[75] This leads to injurer exercising precaution and prevent future harms. This also prevents other potential injurers to exercise precaution. *Reaching the efficient level requires compensation at the margin based on expected societal harm at the margin.*[76]

In case of negligence or fault-based liability, as mentioned earlier, due to unique features of AI: complexity, opacity, multiple parties, self-learning, causality and fault are difficult to prove. This also deters from proving if due care to avoid negligence was performed or not. Therefore, negligence is not an appropriate standard to induce internalization of costs of AI harm.

On the other hand, in case of strict liability, producer of AI will be liable, irrespective of the care taken by the producer. While, it will fully internalize the cost, there is danger of over-internalization. In other words, costs of strict liability can be so high that it

[75]Supra note 61, Cooter and Ulen, p.189

[76] Report on EU Liability Rules for the Age of Artificial Intelligence, Centre on Regulation in Europe, March 2021, p. 39. (EU Liability Report)

prevents the producer from undertaking the activity at all. This can have harmful impact on innovation in highly dynamic field of AI. Another option for the producer, in case of strict liability, is to still perform level care relevant for negligence standard.

Nevertheless, Galasso and Luo claim that strict liability instead of having chilling effect on innovation, can "*encourage firms to develop risk-mitigating technologies and improve the design of their products to reduce the likelihood of harm, and in turn, increasing user trust and take-up."*[77]

At the same time, discovery of harm or fault is still an issue. Thus, even in case of strict liability, there is hurdle in internalization of costs of AI harms. The probability of detection of fault is important in such case. When probability of detection is low, even high compensation amount can be less than the social costs.

It must be noted that in case of fault liability, information costs are much higher and probability of successful prosecution is much lower, as compared to no-fault liability. Also, court has to undertake lesser information costs and enforcement costs because it does not have to ascertain level of care taken by the AI provider/producer.

Thus, strict liability is more suited to AI harms, then negligence. Nevertheless, strict liability suffers from the following shortcomings:

1. It will affect the level of activity and can suppress innovation in AI
2. High compensation amounts can lead to judgement-proof problem wherein resource poor injurer cannot pay the amount of compensation. This is particularly relevant for small and medium innovators.
3. AI harm can be called idiosyncratic risk wherein "from an individual perspective, the damage remains a random event but for the liable party, the outcome is rather predictable."[78]

[77] Supra note 12, Galasso and Luo.
[78]Supra note 76, EU Liability Report, p. 43.

4.4. Insurance

One way of addressing these issues is through insurance. Insurance adds additional paradigm to this scenario. The insured party transfers its risks to the Insurer. This can be called externalization of risk. Bertolini describes it as *"the legal risk of being called to compensate damages is typically dealt with insurance contracts that transform ex post incertitude into an ex-ante – known, thence manageable – cost*."[79]

For instance, if an autonomous car causes accident injuring the driver and third party, then whatever compensation is to be paid to the driver and the third party as per the court orders, will be paid by the insurer. In turn, through subrogation of victim's rights to the insurer, now insurer can claim the amount paid from the insured or distribute the cost throughout the insurance pool, as per the terms of the insurance contract.

It is worth noticing that all the features of AI, which make apportionment of liability in contract or tort liability, difficult are also applicable in the case of AI. The same issues plague insurance contracts. Insurance works on the basis of rule of large numbers. Insurance companies set premiums for the insured which generate revenue. Premium collected from each insured party creates a pool of fund which is used to pay off insured party's liability in case it arises. In order to profitably maintain the pool and pay off the liability, premium amount is important. The premium amount corresponds to the level of risk and probability of liability for the insured activity. In words of Bertolini, premium is function of the risk. [80]Its amount is dependent on the *"likelihood of its occurrence, and the severity of the consequences that may arise once it materializes."*[81]

Therefore, in order to enter into an insurance contract, it is pertinent to have information on likelihood of occurrence and severity.

---

[79] Andrea Bertolini, *Insurance and Risk Management for Robotic Devices: Identifying the Problems*, Global Jurist 16 (3), 2016, 291-314, p. 292

[80] Supra note 79, Bertolini, p.292.

[81] Supra note 79, Bertolini, p.293.

This complicates circumstances for insurance of AI systems. AI features discussed earlier make it difficult to assess the risk with accuracy and in turn, unable to set a premium which reflects the true risks associated with the AI systems. **First**, there is ambiguity regarding legal liability of the parties involved in AI induced harms. AI systems involve multiple players and opacity of system prevents from finding out true role played by each party in the harm caused. **Second**, AI is still a developing field. Not much statistical data (testing data or case laws) is available in the market. If AI is in completely new sector, for example, AI drone delivery devices, then no data is available to evaluate the level of risks. If AI is used in already existing sector, then also available data may not be relevant. Such as, in transport sector, available data is for non-autonomous cars. It is not useful for testing the performance and risks of autonomous cars. **Third**, self-learning nature of AI systems make its uses and risks unforeseeable. Same AI device can be used to play chess and identify cancerous tumors. Bertolini states that, *"this latter problem then intertwines with the empirical issue of determining what kind of harm single devices may bring about. The innovative nature of the application, the multiplicity of possible uses and scenarios in which they could be deployed, raises material concerns with respect to the foreseeability of the – kinds and entity – of damages that could be caused, together with the likelihood of their occurrence".*[82] **Fourth,** it is expected that AI losses might bot be independent and this is essential prerequisite for insurability.[83]

[82] Supra note 79, Bertolini, p.293.
[83] Martin Eling, *How insurance can mitigate AI risks*, 7 November 2019, Available at: https://www.brookings.edu/research/how-insurance-can-mitigate-ai-risks/ (Last accessed on 4 August 2021).

These reasons, Bertolini claims that, "*threaten(s) the very possibility of determining the specific legal risk – probability of being held liable under given circumstances – that each involved party is facing, ultimately prevents its management and insuring."*[84]

Nevertheless, if it is assumed that premium can be estimated accurately for AI systems, then also, issues persist. We will see, how insurance interacts with tort liability and whether it provides appropriate risk-management for AI systems. It can have both positive and negative effect on economic efficiency. Insurer can be considered Superior Risk Bearer who can pool the risks and lead to overall reduction in accident costs. Also, if insurance company is involved, enforcement cost of tort liability laws is reduced. Parties will not invest lot of resources in defending in their case as their risks are already covered through insurance. However, in order to avail these benefits of insurance, the negative effect must be countered. Here contractual incentives can be useful. Contract allow efficient allocation of risks and reduce costs. When injurer or victim enters into a contract with insurer it allocates its risks to insurer. The insurance company with its expertise and economies of scale can reduce the overall costs of risks.

Issues with Insurance:

First, there is problem of **moral hazard**. "Moral hazard arises when the behavior of the insured person or entity changes after the purchase of insurance so that the probability of loss or the size of the loss increases."85 If liability cost of Injurer is born by insurer company, then Injurer does not have any incentive to exercise duty of care and adhere to optimal level of precaution. It is an issue of moral hazard where consequence of Injurer's careless actions does not fall on him/her. Thus, there may be increase in accidents and related social costs. Similarly, in case of Negligence rule or Strict

[84] Supra note 79, Bertolini, p.292.
[85] Supra note 61 Cooter and Ulen, p.48.

Liability with Contributory Negligence, if Victim is insured, the risk will transfer to insurer and Victim will not have any incentive to take precautions.

Second, there is issue of **adverse selection.** *"This arises because of the high cost to insurers of accurately distinguishing between high- and low-risk insurees."*[86] In case of AI systems where insurer cannot accurately estimate the level of risks, it may result in high premiums. That can cause adverse selection of high-risk AI systems as opposed to randomized selection of both high and low risk AI systems. In these circumstances, low-risk AI systems will opt out of the insurance market because they would mot want to pay high premium for their low liability costs. This leads to a vicious cycle where only bad risks are left with the insurer.

Insurance contracts can be sued to control the problem of moral hazard through contractual terms. It can- create exception for grossly negligent acts; allocate fixed amount of accidental losses to insured; allocate fixed percentage of accidental losses to the insured in for of co-insurance and adjust premium amount as per risk level of insured party's activity. In order to get lower premium or avoid other obligations under insurance contract, the insured will choose optimal level of activity and maintain deterrence effect of tort law. This also reduces monitoring cost on behalf of the insurer. Unfortunately, these measures are effective up to a limit.

Additionally, insurer can impose Ex-ante safety standards on inured parties through contractual obligations. For instance, quarterly service and safety check of car to have insurance company bear the risk of accident loss. This is a form of regulation by insurance company. It will ensure that insurer follows optimal level of precaution and fulfil the objective of tort law, i.e., reduction of cost arising from accidents.

However, drawing up of insurance contract for AI systems is difficult, as explained earlier. Here, attempt of private companies and governments to define principles of AI

[86] Supra note 61, Cooter and Ulen, p.48.

safety and accountability play an important rule. These rules, not only, provide rough guidance to the insurer regarding calculation of premium, they can bring more certainty to the otherwise uncertain market. Since, all stakeholders, discussed in Section 3, put emphasis on testing, transparency and explainability, it will gradually bring more certainty to the understanding of inner workings of AI systems. Consequently, insurer will be able to fix a premium which reflects the level of risk involved in the AI system.

Therefore, ex-ante regulation at this stage can, while providing ample space for the growth of innovative AI system, bring legal certainty and facilitate insurance contracts. Like, China and EU proposal provide for sandboxes for development and testing of innovative AI systems. This can provide safe space for dynamic economic efficiency. In fact, a White Paper by Insurance industry and sponsored by Microsoft called for a rules for responsible AI and data science.[87]

But, right now we are far from such legal certainty. Meanwhile, there is possibility that insurers abstain from entering into insurance contract. Then, public authorities can mandate compulsory insurance. However, it is not merely an issue for profitability. Given the grave uncertainty surrounding risks of AI systems, compulsory insurance can result into moral hazard and adverse selection problems. *Absent adequate of statistically relevant data, risks might be hard to define and assess, and a generalized duty to insure might have a strong technology chilling effect*.[88]

Another option is compensation funds. These funds, generally do not fully internalize the costs of an injury or damage. Nevertheless, it can provide some financial support. This can be crucial for maintaining the trust of users in AI systems.

---

[87] Jim DeMarco et al., *We need rules of the road for responsible AI and data science*, Available at: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RWFKCm (Last accessed on 4 August 2021).

[88] *Artificial Intelligence and Civil Liability*, Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies, July 2020.

The insurance industry can play a crucial role in modern digital economy through detection and evaluation of risks[89] The EU Parliament had recommended for *"system of registration for specific categories of advanced robots and adopting a future legislative instrument that should be based either on strict liability or on a risk management approach, in each case supplemented by an obligatory insurance scheme backed up by a fund to ensure that reparation can be made for damages in cases where no insurance cover exists.*"[90] Bertolini suggests that *"to make higher risks more manageable, different approaches might be used, including first- or third-party compulsory insurance, automatic no-fault compensation funds, and liability caps either on their own or in combination with one another."*[91]

Thus, although insurance can solve certain issues associated with the AI system tort liability, it too suffers from similar limitations. Therefore, at present, it is not a viable option due to prerequisites of insurability. However, in future, it can be viewed as a viable and optimal solution to induce optimum level of precaution. Another crucial point is that, even if no regulation is finalized or policy papers by various private and public bodies remain ambiguous, insurance industry has a lucrative market in the form of SI systems. In simpler terms, insurance industry has stake in bringing certainty to the AI systems. This will enable them to insure the vast number of new users and uses which were previously uninsured.

Regulation vs Ex-post Liability:

Shavell notes that administrators and courts have different capabilities when it comes to information.[92] He opines that with respect to technical knowledge, administrators are

[89] Supra note 83, Eling,
[90] European Parliament, *Civil Law Rules on Robotics resolution*, 16 February 2017, Sec.2, 53, 57, 58.
[91] Supra note 79, Bertolini, p.312.
[92] Steven Shavell, *Liability for Harm versus Regulation of Safety*, The Journal of Legal Studies, 13(2), June 1984, 357-374, p. 359.

better placed to collect information, in contradistinction of the courts who excel more in collection of general information.[93] It is also stated that, if liability law and safety regulation impose same standard, then potential defendant conforms to both. This way he/she avoids any ex-ante as well as ex-post liability.[94] Thus, in case of AI systems, administrators are better placed to gather technical information as opposed to the courts. This brings down the transaction costs of enforcement for regulation and makes it more attractive option.

5. Conclusion

In the previous sections, we have seen that AI systems due to their unique characteristics of self-learning or autonomy, have upended the traditional approach to product harms and accountability for those harms. In response to these challenges put forward by the AI systems, many stakeholders have consolidated AI principles. The key principles related to product harm arising from the AI systems are transparency, explainability, testing and validation (conformity assessment), risk-control, human control of AI systems. EU in form of a regulatory proposal has put forward most extensive document. This proposal seeks to establish EU wide harmonized safety standards and ecosystem for the AI systems.

The EU proposal outrightly excludes certain AIs and imposes rigorous obligations on AI s considered high-risk. This is motivated by desire to keep the harms of AI systems to the minimum. However, it can result in adverse effect. Along with tremendous evolution of AI capability, there is also evolution in approach to AI designing. For instance, Stuart Russell recommends that future AI designing should be modified. He states that instead of having predetermined objective for the AI at the time of designing,

[93] Ibid.

[94] Charles Kolstad et al., *Ex Post Liability for Harm vs. Ex Ante Safety Regulation: Substitutes or Complements?*,American Economic Review, 80, 1990, 888-901.

the objectives should change during the course of use of AI.[95] Keeping this in mind, any regulation should be adaptable to changing technology. With the progress of technological development, transaction costs of private bargaining and prosecution of a tort liability case also alter. Therefore, any regulation should have sufficient flexibility by being technologically neutral.

Moreover, there should be scope for regular and timely review of a regulation to keep it up to date with real world technological advancement and legal challenges. The EU proposal (Article 84) provides for annual review of the high-risk AI systems and once in three years review of the whole regulation. This is an innovative regulatory instrument to keep the regulation up to date with times.

We also looked at contractual and tort liability in light of AI systems self-learning capabilities. Contractual as well as Tort (negligence and strict) liability are not perfect to internalize all the externalities of an AI system. Due to involvement of multiple parties and opacity of the AI system, enforcement of contract includes high transaction costs. Similarly, fault liability does not mitigate all the risks and has high transaction costs for successful prosecution. Strict liability has lesser transaction costs, but can cause a chilling effect on innovation. Furthermore, it also suffers from judgement proof-problem. Insurance, is also difficult at this stage, because AI systems do not fulfill all the prerequisites of insurability as of today.

In the light of all these considerations, a hybrid approach is best for the AI systems.

Regulation or self-regulation with minimum safety directions and supported by regulatory sandboxes can minimize the risks ex-ante while promoting innovation through experimentation. These regulations or policies by private entities, provide complementary support for tort liability as well as insurability. These policies even at this stage present a blueprint which provides guidance to the courts of how to measure

[95] Supra note 17, Russell and Norvig, p.1020

level of care (in absence of any specified liability standard) and insurers in ascertaining the risks of AI systems.

Thus, a hybrid approach of strict liability coupled with ex-ante regulation for safety and ex-post insurance cover would be best to address the AI harms, until technological progress changes the parameters of risk. It can also be called a 'smart mix'[96] It is explained that "*the idea behind smart regulation is that various regulatory and governance instruments, both public and private and both international and local, can be combined into mixes of complementary instruments and actors, tailored to the specific needs of the situation."*[97] This reasoning can be extended to AI systems as well. They are multifaceted with multiple challenges. Such challenge requires that a "smart mix" instruments works in tandem to holistically address the accountability and safety issues caused by the AI systems.

In conclusion, challenges of AI systems are not going to be resolved anytime soon. An open, flexible and innovative approach to legal challenges is best answer for now.

******

[96] *The Concept of Smart Mixes for Transboundary Environmental Harm* in *Smart Mixes for Transboundary Environmental Harm*, Judith van Erp, Michael Faure, Nollkaemper André, and Niels Philipsen, eds, Cambridge Studies on Environment, Energy and Natural Resources Governance, Cambridge University Press, 2019, p.5.
[97] Ibid.

Bibliography

2018 in Review: 10 AI Failures, 11 December 2018, Available at: https://medium.com/syncedreview/2018-in-review-10-ai-failures-c18faadf5983 (Last accessed on 4 August 2021).

Alberto Galasso and Hong Luo,*When does product liability risk chill innovation? evidence from medical implants,* National Bureau of Economic Research, No. wp 25068, September 2018.

Anat Lior, *The AI Accident Network: Artificial Intelligence Liability Meets Network Theory,* Tulane Law Review 95, 2020.

Andrea Bertolini, *Insurance and Risk Management for Robotic Devices: Identifying the Problems,* Global Jurist 16 (3), 2016, 291-314, p. 292

Artificial Intelligence and Civil Liability, Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies, July 2020.

Benjamin Cheatham, Kia Javanmardian, and Hamid Samandari, *Confronting the risks of artificial intelligence*, 26 April, 2019, McKinsey Global Institute, Available at: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-

insights/confronting-the-risks-of-artificial-intelligence (Last accessed on 4 August 2021).

Brent D. Mittelstadt, et al., The Ethics of Algorithms: Mapping the Debate, Big Data & Society, December 2016.

Bryant W. Smith, *Automated Driving and Product Liability,* Michigan State Law Review Journal, 1, 2017, 1-74.

Charles Kolstad et al., *Ex Post Liability for Harm vs. Ex Ante Safety Regulation: Substitutes or Complements?,* American Economic Review, 80, 1990, 888-901.

Chinese National Governance Committee for the New Generation Artificial Intelligence, led by China's Ministry of Science and Technology, *Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence*, 2019, Available at: http://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html (Last accessed on 4 August 2021).

Chris Anderson, *The End of Theory,* Wired , July 2008.

Christaine Wendehorst, *Sale of Goods in the Digital Age – From Bipolar to Multi-party Relationships* in The Age of Uniform Law. Essays in honour of Michael Joachim Bonell to celebrate his 70th birthday, UNIDROIT, 2016, 1873–1887.

Cynthia Rudin and Joanna Radin, *Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition*, Harvard Data Science Review, 1(2) 2019.

Dario Amodei, et al., *Concrete problems in AI safety*, 2016, Available at: *arXiv preprint arXiv:1606.06565* (Last accessed on 4 August 2021), pp.2-3

Darrell M. West and John R. Allen, *How artificial intelligence is transforming the world*, Brookings, 24 April 2018, Available at: https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/ (Last accessed on 4 August 2021).

Elena A. Kirillova, *Artificial Intelligence as a New Category of Civil Law*, Journal of Advanced Research in Law and Economics, 11(1), 2020.

European Commission, Proposal for a Regulation of the European Parliament and the Council Laying Down Harmonised Rules of Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 2021/0106, 21 April, 2021 (EU Proposal), Article 3.1.

European Commission, *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, 19 February 2020. (EU Safety Report)

European Parliament, *Civil Law Rules on Robotics resolution*, 16 February 2017, Sec.2, 53, 57, 58.

F. Patrick Hubbard, *Sophisticated Robots": Balancing Liability, Regulation, and Innovation,* Florida Law Review, 66(5), 2015.

Google, *AI at Google: Our Principles,* 2018, Available at: https://www.blog.google/technology/ai/ai-principles/ (Last accessed on 4 August 2021).

Guido Calabresi, *The Costs of Accidents*, 1970.

Herbert Dawid and Gert Muehlheusser, *Smart Products: Liability, Timing of Market Introduction and Investments in Product Safety*, Bielefeld University, 2019.

IBM, *IBM Everyday Ethics for AI*, 2019, Available at: https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf (Last accessed on 4 August 2021).

Jacques Bughin, et. al, *Notes from the AI frontier: Modelling the impact of AI on the world economy*, McKinsey Global Institute, 4 September 2018, Available at: https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy (Last accessed on 4 August 2021).

Jessica Fjeld et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Berkman Klein Center Research Publication No. 2020-1, January 15, 2020, Available at SSRN: https://ssrn.com/abstract=3518482 or http://dx.doi.org/10.2139/ssrn.3518482 (Last accessed on 4 August 2021).

Jim DeMarco et al., *We need rules of the road for responsible AI and data science*, Available at: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RWFKCm (Last accessed on 4 August 2021).

John McCarthy, *What is Artificial Intelligence?* 2007, Available at www.formal.stanford.edu/jmc/whatisai.pdf (Last accessed on 4 August 2021).

John Villasenor, *Products liability law as a way to address AI harms*, Brookings, 31 October 2019, Available at: https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms/ (Last accessed on 4 August 2021).

Luke Dormehl, Thinking Machines: The Quest for Artificial Intelligence and Where It's Taking Us Next, Tarcher Perigee, 2017, p.8.

Martin Ebers, *Regulating AI and Robotics* in *Algorithms and Law*, Cambridge University Press, 2020, 37-99, p.41.

Martin Eling, *How insurance can mitigate AI risks*, 7 November 2019, Available at: https://www.brookings.edu/research/how-insurance-can-mitigate-ai-risks/ (Last accessed on 4 August 2021).

Matthew Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies and Strategies*, Harvard Journal of Law and Technology, 29(2), 2016, 354-400, pp. 363–364.

Matthias Leese, *The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-discriminatory Safeguards in the European Union*, Security Dialogue, 45(5), 2014, 494-511.

Michael Li, *Another Self-Driving Car Accident, Another AI Development Lesson*, 16 November 2019, Available at: https://towardsdatascience.com/another-self-driving-car-accident-another-ai-development-lesson-b2ce3dbb4444 (Last accessed on 4 August 2021).

Microsoft, *AI Principles*, 2018, Available at: https://www.microsoft.com/en-us/ai/our-approach-to-ai (Last accessed on 4 August 2021).

Niti Aayog, *National Strategy for Artificial Intelligence: #AI for All, 2018,* Available at: https://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf (Last accessed on 4 August 2021).

Olivia Solon, *Roomba Creator responds to reports of 'poopocalpse': We see this a lot*, 15 August 2016, Available at: https://www.theguardian.com/technology/2016/aug/15/roomba-robot-vacuum-poopocalypse-facebook-post (Last accessed on 4 August 2021).

Product Liability Directive 85/374/EEC, 1985.

Report on EU Liability Rules for the Age of Artificial Intelligence, Centre on Regulation in Europe, March 2021, p. 39.

Rob Kitchin, *Thinking Critically about and Researching Algorithms*, Information, Communication and Society 20(1) 2017, 1–14.

Robert Cooter and Thomas Ulen, *Law and Economics,* 2016, p.281.

Ronald H. Coase, *The Problem of Social Cost*, Journal of Law & Economics 1(3) 1960.

Steven Shavell, *Liability for Harm versus Regulation of Safety*, The Journal of Legal Studies, 13(2), June 1984, 357-374, p. 359.

Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th edn., Pearson, 2021, p.1.

Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*, Viking, 2019, p.11.

Tencent Institute, *Six Principles of AI*, 2017, Available at: http://www.kejilie.com/iyiou/article/ZRZFn2.html (Last accessed on 4 August 2021).

The Concept of Smart Mixes for Transboundary Environmental Harm in *Smart Mixes for Transboundary Environmental Harm*, Judith van Erp, Michael Faure, Nollkaemper André, and Niels Philipsen, eds, Cambridge Studies on Environment, Energy and Natural Resources Governance, Cambridge University Press, 2019, p.5.

Will Knight, *Defeated Chess Champ Garry Kasparov Has Made Peace With A*I, 2 February, 2021, Available at: https://www.wired.com/story/defeated-chess-champ-garry-kasparov-made-peace-ai/ (Last accessed on 4 August 2021).