University of St Andrews
Law Journal
Institute of Legal and Constitutional Research

AI SYSTEMS AND LIABILITY: AN ASSESSMENT OF THE APPLICABILITY OF STRICT LIABILITY & A CASE FOR LIMITED LEGAL PERSONHOOD FOR AI

by Louisa McDonald

I.   Introduction: Machine Learning Algorithms and Product Liability

Under a classic conception of product liability, if a defective product causes damage to persons or property, the manufacturers and other persons involved in creating the product are liable for the damage caused[1]. If the plaintiff is able to prove that the damage was caused by the product, then those involved in producing the product are strictly liable and owe the victim compensation[2]. This is the position of EU legislation on liability for defective products, formalised by the EU Product Liability Directive [1985], a defective product causes any physical damage to consumers or their property, the producer has to provide compensation irrespective of whether there is negligence or fault on their part[3].

Even in cases of strict (rather than fault-based) liability for products, a causal connection between the producers or operators and the damage caused is assumed. However, this assumption becomes more problematic in the case of recent AI technologies, which exhibit a degree of autonomy that may mean that they are able to perform acts – including acts in the law – which the human agents involved (programmers, manufacturers, operators, etc.) could not possibly foresee.

Previously, electronic agents could be simply regarded as tools and the correlative legal issues that arose from their usage could be entirely attributed to human agents. The Uniform Electronic Transactions Act (UETA) [1999][4] recognises electronic agents as being limited to a

---

[1] Jonathan Law and Elizabeth A. Martin, "Product Liability" in A Dictionary of Law (7th Edition), Online Version, (Oxford, 2014)

[2] Law and Martin, "Product Liability"

[3] European Commission, *Council Directive 85/374/EEC of July 25 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products (Product Liability Directive)* OJ L210/29 (1985)

[4] National Conference of Commissioners on Uniform State Laws, "Uniform Electronic Transactions Act (UETA)" (1999) 1/20/00

'tool' function[5] . However, recent AI systems can no longer be classed as mere 'tools' in this way because of the use of machine learning algorithms, which lead to the AI system having a degree of autonomy.

Machine learning enables systems to learn and improve from experience[6]. Machine learning algorithms work by employing artificial neural networks: these are simplified models of the brain composed of large numbers of units together with weights which measure the strength of connections between the units[7]. A simple neural network has an input, hidden, and output layer. Deep neural networks – the kind that enable machine learning – have more than one hidden layer[8]. Such deep neural networks exhibit machine learning capacities which can move significantly beyond the original programming. For example, a Deep Q-Network program used reinforcement learning to learn to play Atari 2600 games with no prior knowledge, discovering strategies not known to its programmers[9].

The autonomous AI of today can perform acts which bring about legal consequences. This is especially pertinent in the stock market with the increasing prevalence of trading bots, some of which, such as 'B-Cube AI', can perform trades autonomously[10], so that operators might not even be aware of the trades[11]. The UETA does not cover such algorithmic contracts, because machine learning algorithms are not programmed by people and therefore fall outside of its scope[12].

If AI systems can engage in activities such as producing contracts, then it appears that they can perform acts in the law: they can "produce legal rights and obligations through their acts and actions"[13]. However, since AI is not currently granted legal personhood, they do not have legal capacity and cannot be party to a legal transaction[14].

---

[5] Pınar Çağlayan Aksoy, "AI as Agents: Agency Law." in *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, eds. Larry A. Di Matteo, Cristina Poncibo and Michel Cannarsa (Cambridge University Press, 2022): 147

[6] Expert.AI, "Expert.AI: What is Machine Learning? A Definition" (March 2022) https://www.expert.ai/blog/machine-learning-definition/

[7] Cameron Buckner and James Garson, "Connectionism: The Stanford Encyclopedia of Philosophy" (2019) https://plato.stanford.edu/archives/fall2019/entries/connectionism/

[8] Buckner and Garson, "Connectionism"

[9] DeepMind Technologies, "Playing Atari with Deep Reinforcement Learning." (2015) https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf.

[10] B-Cube-AI, "Cutting-Edge AI Crypto Trading Bots" (Accessed 2023) https://b-cube.ai/

[11] Aksoy, "AI as Agents: Agency Law": 146

[12] *Ibid.,* 148.

[13] Aksoy, "AI as Agents: Agency Law", 147

[14] *Ibid.*

We are therefore faced with a dilemma for AI liability. Either certain kinds of autonomous AI are granted legal personhood, and therefore are legal agents and can be liable for their actions, or the liability for the actions of an AI system rests entirely on the human agents which contributed to its production and operation, even if the actions of the AI could not possibly have been foreseen by humans. Fenwick and Wrbka propose two potential models for AI liability:

Personhood model: victims are granted compensation directly from the AI: AI is accorded legal personhood.

Liability model: AI is not accorded independent legal personhood and victims receive compensation from some other legal person (human person or company)[15]

In this essay, I will ultimately defend the personhood model, on the grounds that the liability model confers unfair consequences both on victims of the actions of AI and on those responsible for producing the AI. I will begin by giving a definition of legal personhood as a cluster concept which can comprise more than just natural persons, following Kurki's analysis of legal personhood as a cluster of passive and active incidents. I will then explain the EU's AI Liability Directive[16] and present the arguments given by the Commission as to why a liability model is apt. Next, I will counter the Commission's claim by presenting two pressing problems for liability models, before making a positive case for the personhood model. Finally, I will counter some common objections to AI personhood.

II. Legal Personhood: The Cluster Concept View

i) The Nature of Legal Personhood

It is important to distinguish a legal person from a natural person: rather than being defined by any kind of ontological considerations, a legal person is simply an entity which is treated as a person by the law and has the relevant rights and capacities. The Oxford Dictionary of Law defines

---

[15] M. Fenwick and S. Wrbka, "AI and Legal Personhood." In *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, eds. Larry A. Di Matteo, Cristina Poncibo and Michel Cannarsa (Cambridge University Press, 2022): 288

[16] European Commission, *Proposal for a Directive of the European Parliament and of the Council 2022/0303 of 28 September 2022 on adapting non-contractual civil liability rule to artificial intelligence (AI Liability Directive)* OJ C496, (2022)

a legal person as either "a natural person (i.e. a human being) or a juristic person"[17] where a juristic person is "an entity, such as a corporation, that is recognized as having legal personality, i.e. it is capable of enjoying and being subject to legal rights and duties."[18] If AI were to be granted legal personhood, it would fall under the category of a juristic person.

Visa Kurki understands legal personhood as an institutional fact. This term is taken from Searle, who used it to refer to facts in the world which depend on human institutions but are nonetheless objective facts, such as the fact of being a US citizen[19]. Kurki observes that "a necessary condition for the legal personhood of any entity is that the entity is treated as a legal person by the prevailing legal system"[20]. Whether an entity is granted legal personhood is determined by the actions of legal institutions, such as in the landmark US case Citizens United v. FEC [2010][21], which set the precedent for corporations being granted free speech rights in the same way as human persons.

ii)   Legal Personhood as a Cluster Concept

Kurki also conceives of legal personhood as a cluster concept[22]: a cluster concept is comprised of a weighted list of criteria, such that none of these criteria alone is either necessary or sufficient for membership. A famous example given by Wittgenstein is the concept 'game': our uses of the word vary so much that no one unifying quality can pick out everything denoted by the word 'game', and therefore we should instead think of it as a cluster concept which varies based on each instance[23]. Legal personhood can be seen as a cluster concept because it consists in a cluster of rights and responsibilities which vary based on the type of legal person in question. For example, a child may be subject to a number of rights but may not possess certain legal competences until they reach a certain age, e.g., they cannot vote until they reach the age of eighteen.

---

[17] Jonathan Law and Elizabeth A. Martin, "Legal Person" in *A Dictionary of Law (7th Edition), Online Version*, (Oxford, 2014), https://www.oxfordreference.com/display/10.1093/acref/9780199551248.001.0001/acref-9780199551248;jsessionid=3CA80E5E8BCEF2DD0614AC49860DB61F

[18] Jonathan Law and Elizabeth A. Martin, "Juristic Person" in *A Dictionary of Law (7th Edition), Online Version*, (Oxford, 2014)

[19] John R. Searle, *The Construction of Social Reality* (Penguin, 1995): 1

[20] Visa A.J Kurki, *A Theory of Legal Personhood* (Oxford Academic, 2019): 92

[21] *Citizens United v. Federal Election Commission* [2010] *558 U.S. 310 (2010)*

[22] Kurki, *A Theory of Legal Personhood*: 93

[23] Ludwig Wittgenstein, *Philosophical Investigations*, eds. and trans. P.M.S. Hacker and Joachim Schulte (Oxford: Wiley-Blackwell, 1953, ed. 2009): 65

Kurki terms the constituent components of the cluster property 'incidents', meaning non-procedural claim-rights and liabilities that can be held or acquired by an entity[24]. He divides up these incidents into passive and active incidents[25]. Roughly speaking, to possess a passive incident of legal personhood is to be able to be subject to rights and legal protections[26]. Active incidents, on the other hand, concern being able to administer legal competences (e.g., enter into a contract) and what Kurki terms 'onerous legal personhood' (having legal responsibilities in tort and criminal law)[27].

iii)   Dependent and Independent Personhood

An implication of the cluster concept view is that it allows for the passive and active incidents to vary based on the legal person in question: no one passive or active incident will be individually necessary or sufficient to constitute personhood. Importantly, the cluster concept view allows for legal persons to be both dependent and independent.

Whilst an independent legal person can exercise some or all its rights through its own agency, a dependent legal person can only act through the agency of another legal person in exercising some or all its rights[28]. It is important to note that dependent legal persons can still consist of active legal positions and can be duty-bearers as well as rights-bearers[29]. This is a plausible way of characterising the type of legal personhood that corporations have, since they are dependent on the agency of their constituent members, but nonetheless can be subject to both passive and active incidents of legal personhood. I would equally argue that if AI systems were given legal personhood, it would be a kind of dependent legal personhood, because they are dependent on the agency of programmers and manufacturers in order to exercise their legal capacities.

The status of dependent legal personhood is not undermined by the AI systems having autonomy. An analogy can be made with corporations. Groups can have agency separate from the sum of that

---

[24] Kurki, *A Theory of Legal Personhood*: 95
[25] *Ibid.*
[26] *Ibid.*
[27] *Ibid.,* 96
[28] Claudio Novelli, "Legal Personhood for the Integration of AI Systems in the Social Context: A Study Hypothesis." in *AI & Society* 1, (2022): 6
[29] *Ibid.,* 7

of their members, even though their agency depends on that of their individual members. Consider the following example:

A hiring committee is looking for a candidate who fulfils all the following criteria: X, Y and Z. Three people are on the committee. In each instance, if the majority believe that the candidate fulfils the relevant criterion, the verdict will be a 'yes':

Panelist	(i)	(ii)	(iii)	Verdict
X	Yes	Yes	No	Yes
Y	Yes	No	Yes	Yes
Z	No	Yes	Yes	Yes

Note that none of the constituent members individually believe that the candidate fulfils all the relevant criteria. However, the group verdict is that he does. We can thereby see how group agency can come apart from individual agency, even though it depends on the agency of individuals. Similarly, although the agency of AI systems might depend on its creators, it can come apart from it.

III.   The 2022 EU AI Liability Directive: A Liability Model

The European Commission acknowledges the challenges that come with autonomous AI systems when it comes to personal responsibility. The EU website for the Regulatory Framework Proposal on Artificial Intelligence states that, "it is often not possible to find out why an AI system has made a decision or prediction and taken a particular action. So, it may become difficult to assess whether someone has been unfairly disadvantaged, such as in a hiring decision or in an application for a public benefit scheme"[30].

---

[30] European Commission, "Regulatory Framework Proposal on Artificial Intelligence" (2021) https://commission.europa.eu/document/f9ac0daf-baa3-4371-a760-810414ce4823_en

The EU's AI Liability Directive (2022)[31] attempts to deal with the problem of autonomous AI systems causing harm by introducing a framework of strict rather than fault-based liability for developers, producers, and users of AI technology. The Directive mainly builds on and adapts the Product Liability Directive (1985)[32]. It roughly consists of: (i) measures to ease the burden of proof for victims trying to prove their liability claim, and (ii) a review mechanism to re-assess, in particular, the need for harmonising strict liability for AI use cases with a particular risk profile[33].

(i)   Easing the burden of proof: the 'presumption of causality'

The argument for (i) is rooted in the acknowledgement that general fault-based liability, whilst appropriate for other kinds of product liability, is defective in the case of autonomous AI. It typically requires the person to prove a negligent or intentionally damaging act or omission by the person potentially liable for that damage, and finding this proof can be more complicated in the case of AI systems since, if the system acts autonomously, no one identifiable person has caused the damage[34]. The status quo is that it would be excessively difficult for victims to meet the burden of proof, and therefore the burden of proof ought to be made lighter. The Directive aims to do this by introducing a 'presumption of causality', defined by the Commission as follows: "if victims can show that someone was at fault for not complying with a certain obligation relevant to the harm, and that a causal link with the AI performance is reasonably likely, the court can presume that this non-compliance caused the damage."[35]. Note that the 'presumption' still requires proof of some party not complying with an obligation, and proof of a causal link between the AI and the harm.

(ii) Encouraging progress: harmonising liability laws

The motivation for (ii) stems from the observation that AI liability laws are not harmonised across the national civil liability laws of EU member states, which is likely to hinder overall technological

---

[31] European Commission, *AI Liability Directive*
[32] European Commission, *Product Liability Directive*
[33] European Commission, *AI Liability Directive*: 14-16
[34] *Ibid.,* 13
[35] *Ibid.*

progress, since it is difficult for businesses and developers to anticipate how liability rules will be applied to their software[36]. Since the EU Artificial Intelligence Act (2021) already provides risk categories for AI[37], AI liability laws should be harmonised across the EU according to these risk categories. The 'presumption of causality' is applied in various ways according to the level of risk of the AI concerned, for example, in the case of non-high-risk AI systems, the court must first determine that it is excessively difficult for the claimant to produce the causal link[38]. By creating a unified framework of liability legislation that operates across the EU, the Directive therefore claims that it will improve conditions for developers of AI systems "by preventing fragmentation and increasing legal certainty through harmonised measures at EU level, compared to possible adaptations of liability rules at national level."[39]

The legislation is therefore supposed to be beneficial both to victims seeking compensation for harm caused by AI systems and for developers seeking to produce AI systems in accordance with liability laws. In the next section, I will call both assumptions into question.

IV.   Problems with the Liability Model

The AI Liability Directive is correct to observe that fault-based liability schemes are not appropriate for autonomous AI. However, I would further argue that Directive's liability model goes wrong in assuming that a system of strict liability will be the optimal model for AI liability legislation. This is because a system of strict liability is likely to have undesirable consequences both for those involved in producing AI systems and for those harmed by AI systems seeking compensation. It is thereby unlikely to achieve the Commission's main aims: it is unlikely to encourage progress in the development and production of AI systems, and it is unlikely to actually be beneficial for victims seeking compensation.

(i)   Consequences for Progress in the Development of AI Systems

---

[36] *Ibid.,* 17

[37] European Commission, *Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artifical Intelligence Act) and amending certain union legislative acts* 2021/0106 (COD) (2021)

[38] *Ibid.,* 13

[39] *Ibid.,* 4

Whilst the Directive claims that the harmonisation of AI liability laws will make production easier, this alone is not strong enough to encourage production if the laws themselves are a hindrance to innovation and to progress. With a system of strict liability, costs are systematically placed on a single party regardless of fault. This may place unfair constraints on human agents and lead to undesirable consequences for technological progress. Novelli points out that not only would such a system potentially disincentivise production, but they might also "discourage consumer diligence as someone else will always be held liable"[40]; the system of strict liability proposed in the Directive appears to shift the burden almost entirely to those involved in producing the AI system and does not leave much room for responsibility on the part of the consumer, making an apparently arbitrary distinction.

Whilst it is in everyone's interests to prevent the emergence of harmful AI, hindering the progress of AI development on a more general level is not desirable. There are many examples of AI which can transform human life for the better, including the World Bee Project, which uses AI technology to monitor pollinator and biodiversity declines in order to help find long-term solutions to the problem[41]. Another example is Facing Emotions, a project designed by Huawei together with the Polish Blind Association which allows the visually impaired to 'see' emotions on people's faces by translating them into sound[42]. It is important to remember the opportunities that come with the emergence of AI technologies as well as the risks; a liability system that subjects developers to conditions that are overly harsh risks suppressing potentially beneficial AI systems.

(ii) Consequences for Victims of AI Systems Seeking Compensation

Perhaps more significantly, it is not clear that a system of strict liability would have the intended effect of making it easier for those harmed by AI systems to seek compensation.

---

[40] Novelli, "Legal Personhood for the Integration of AI Systems", 4

[41] The World Bee Project, *The World Bee Project: Protecting Pollinators, People & The Planet* (Accessed 2023) https://worldbeeproject.org/

[42] Campaigns of the World, *Campaigns of the World: Huawei Facing Emotions* (Accessed 2023) https://campaignsoftheworld.com/technology/huawei-facing-emotions/

The 'presumption of causality' proposed by the Directive still requires that the defendant show that "someone was at fault for not complying with a certain obligation relevant to the harm, and that a causal link with the AI performance is reasonably likely, the court can presume that this non-compliance caused the damage."[43]. The presumption is therefore applied to the causal link between the non-compliance and the damage caused; proof that someone acted in a non-compliant way (and that this is reasonably likely to be causally linked to the performance of the AI system) is still required. As Novelli points out, proving such matters can be overly onerous for the victim[44]. This is because, in the case of autonomous AI systems, which are produced, used, and developed by multitudinous actors, it can be difficult or even impossible to identify one person that is responsible for the non-compliance. It may even be the case that no individual person can plausibly be identified.

To illustrate this difficulty, an analogy can be made with corporate personhood. In English Common Law, the offence of corporate manslaughter was created to overcome the limitations of the Common Law offence of gross negligence manslaughter when applied to corporations[45]. The offence of corporate manslaughter was created by Section 1 of the Corporate Manslaughter and Corporate Homicide Act (2007)[46]. Previously, under the Common Law, for a corporation to be guilty of such an offence, it was necessary to identify a 'controlling mind', i.e., a senior individual who could be said to embody the company and bear the responsibility for the gross negligence[47]. This created problems when no such individual could be identified. In the case of R v P&O European Ferries (Dover) Ltd, a ferry – the Herald of Free Enterprise – capsized, and 190 passengers were killed as a result[48]. However, manslaughter charges were dismissed since the director of P&O ferries had not appointed anyone to be responsible for health and safety; no 'controlling mind' could be identified[49]. The company ended up getting away with no charges. The 2007 Act aimed to widen the scope of the offence so that the focus shifted to the overall management of the organisation's activities, rather than the actions of individuals[50].

---

[43] European Commission, *Questions & Answers: AI Liability Directive* https://ec.europa.eu/commission/presscorner/detail/en/QANDA_22_5793
[44] Novelli, "Legal Personhood for the Integration of AI Systems": 4
[45] Crown Prosecution Service, *Legal Guidance, Violent Crime: Corporate Manslaughter* (July 2018) https://www.cps.gov.uk/legal-guidance/corporate-manslaughter
[46] *Corporate Homicide Act (2007)* (CMCHAct)
[47] Crown Prosecution Service, *Legal Guidance, Violent Crime: Corporate Manslaughter*
[48] R v P&O Ferries [1991] 93 CAR 72
[49] *Ibid.*
[50] Crown Prosecution Service, *Legal Guidance, Violent Crime: Corporate Manslaughter*

The example of R v P&O Ferries pertains to the criminal rather than civil law, but nonetheless demonstrates how the need to identify a responsible individual can be disadvantageous to victims seeking compensation, because in cases where no such individual can be identified, the outcome could be that the victim simply cannot receive compensation at all. In the case of AI systems, even when a 'presumption of causality' is in place, it may be similarly difficult or even impossible to identify an individual or set of individuals who caused the damage to take place. Furthermore, if the AI system acted autonomously and no individual could possibly have foreseen its actions, then it seems even the causal link is missing between the human agents and the damage caused by the AI system, and therefore it is difficult to see how the victim can claim compensation.

The AI Liability Directive's solution of strict liability therefore does not adequately meet the challenges posed by liability for advanced AI. It risks hindering developmental progress and, more importantly, creating further problems for victims seeking compensation. In the next chapter, I will therefore give a positive consequentialist argument for adopting an alternative model for AI liability: the personhood model

V. A Case for the Personhood Model

An alternative model to the liability model is the personhood model: that is, to grant AI the status of legal persons. Since legal personhood is largely a functional concept that concerns what kind of legal responsibilities and rights an entity can be subject to, it is apt to use a pragmatic-consequentialist line of reasoning. The argument I will advance here is that granting legal personhood to autonomous AI systems could result in legal simplification which would make it easier for injured parties to claim compensation than it would be on the liability model.

i) Legal Simplification and Compensation

Fenwick and Wrbka point out that it has been claimed that the fact that 'someone' is responsible for building the AI system that causes harm is a reason for not granting legal personhood to the

AI[51]. There are several problems with this line of reasoning. First, as has been argued in the previous chapter, it can be exceedingly difficult to trace back the harmful actions of the AI to a specific, easily identifiable person. AI technologies involve multitudinous actors, and it is not clear that any one party can be designated as causally responsible for the harm or non-compliance[52]. Furthermore, the same observation – that there are human actors ultimately causally responsible for the harm occurring – could be made of corporations, yet corporations are granted legal personhood[53]. Moreover, as has been demonstrated in the previous chapter, affording the status of legal personhood to corporation can be beneficial to those seeking compensation because otherwise it would be difficult to identify a responsible individual or responsible individuals. Similarly, granting a situation-specific form of legal personhood to AI may offer greater clarity and, presumably, more opportunity to seek compensation for harm[54].

Arguing similarly on the basis of legal simplification, Novelli argues that the responsible party in the case of AI liability ought to be the AI itself, which would make it far less costly and complicated to identify the liable party[55]. If the AI itself, rather than the human agents involved in its manufacture and deployment, is held to be the liable party, then the type of legal personhood granted to such an AI system must allow for the victim to receive adequate compensation given that the system itself is liable.

ii)   What AI Personhood Would Look Like

If legal personhood is to be conferred on certain kinds of AI systems, it would clearly be a different kind from that conferred on natural persons, since an AI system has different capacities and is not capable of conscious thought.

As has been demonstrated in (I), one important way in which legal personhood can vary based on entities is that it can be dependent or independent, and AI personhood would need to be of the dependent kind: this acknowledges that the AI can only act through the agency of another legal person in exercising some or all of its rights. AI personhood would therefore function in a similar

---

[51] Fenwick and Wrbka, "AI and Legal Personhood": 294
[52] *Ibid.*
[53] *Ibid.*
[54] *Ibid.*
[55] Novelli, "Legal Personhood for the Integration of AI Systems": 5

way to corporate personhood. Importantly, dependent legal personhood of this kind can still consist of active legal positions: dependent legal persons can be duty-bearers as well as right-bearers[56]. Following MacCormick[57], Novelli suggests that the two most important legal positions that AI personhood would need to incorporate would be transitional capacity – the power to enter into and create legally salient relationships – and liability capacity – the susceptibility to legal imputations for civil wrongdoings or criminal offences[58]. This tracks autonomous AI's current capacities, e.g., entering into contracts through trading.

An initial concern about an AI itself having liability capacity is that machines lack assets, and therefore it is difficult to see how they would be able to help victims recover losses. A plausible model of AI personhood would therefore have to either bestow assets on AI through

state-enforced minimum asset requirements or impose some kind of mandatory liability insurance for AI[59]. Both of these solutions recognise the dependent legal status of AI systems. If AI were granted assets with which to compensate victims, these could be the sum of a mandatory contributions from all those involved in producing and deploying the system, and the combined pool of the assets would legally belong to the AI; a minimum requirement for this would "oblige other parties to provide the funds necessary to satisfy potential damages claims"[60]. Alternatively, a policy of mandatory liability insurance for AI would be provided by the other natural and legal persons involved in the production and deployment of the AI system[61]. Both a policy of mandatory insurance and a mandatory minimum asset requirement for AI would ensure that there was a clear way to compensate victims who had successfully sued the AI system for damages.

iii)   Objections to AI Personhood & Responses

AI personhood can be objected to on ontological grounds. Although autonomous AI systems can do some of what humans can do, they lack traits such as consciousness and moral responsibility and therefore cannot be legally responsible in the same sense as natural persons.

---

[56] *Ibid.,* 7
[57] Neil MacCormick, *Institutions of Law: An Essay in Legal Theory* (Oxford University Press, 2007)
[58] Novelli, "Legal Personhood for the Integration of AI Systems: 7
[59] Fenwick and Wrbka, "AI and Legal Personhood": 301-2
[60] *Ibid.,* 302
[61] *Ibid.*

However, this objection misunderstands the nature of legal personhood. Fenwick and Wrbka point out that it is important to disentangle personhood in the legal sense from ontological questions about how similar AI systems are to natural persons[62]. Legal personhood, as we have seen in (II), depends not whether an entity instantiates a given set of qualities, but rather is a status attributed by a particular legal system which varies in its nature depending on what active and passive instances can be attributed to the entity. What counts towards an entity being given legal personhood therefore lies not in its similarity to human persons, but in the extent of its capacities. In the case of AI, it appears that the capacity for autonomy, which includes being able to perform actions that constitute acts-in-the-law, is a key consideration; some degree of autonomous decision making would seem to be a pre-condition of active legal personhood[63]. Novelli points out that autonomous AI systems thereby challenge the traditional legal distinction between things and persons because they are "equipped with an epistemic and practical authority over their behaviour"[64].

There are also instrumentalist objections to legal personhood, the most common being that it could allow for the natural persons involved in AI production to evade responsibility and liability. Furthermore, the evasion of liability could bring about the wrong kind of incentives for those involved in producing and developing AI, since they would no longer have to take such great precautions, knowing that they would not be held liable.

Whilst this objection is more compelling than the ontological objection, I believe the concerns can be adequately addressed by placing the kind of constraints I have described in (ii) on those involved in creating autonomous AI systems: either the provision of assets or mandatory liability insurance. This would translate into a financial burden on those creating the AI which would disincentivise them from negligence in the production or use processes. It is beyond the scope of this essay to establish whether the long-term effects on incentives for producers and developers are better under a personhood model as opposed to a liability model, but I would argue that we have no definitive reason to believe that they would be worse, given the financial constraints the producers and developers would be subject to under a personhood model.

---

[62] *Ibid.,* 292
[63] *Ibid.,* 289
[64] Novelli, "Legal Personhood for the Integration of AI Systems: 5

VI.   Conclusion

I have sought to demonstrate in this essay that a model of strict liability is not appropriate for the kinds of AI systems we are increasingly confronted with, which exhibit a degree of autonomy that means they are not necessarily under the control of the human persons involved in producing and developing them. I have attempted to provide an alternative model for AI liability: granting autonomous AI systems a kind of legal personhood. If AI personhood were to become a reality, it would need to be characterised in a lot more detail than I have gone in to here, but at very least, I have aimed to sketch the beginnings of a model for AI personhood, and to show that it is not an option that ought to be ruled out.

BIBLIOGRAPHY

Aksoy, Pınar Çağlayan. "AI as Agents: Agency Law." in *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, eds. Di Matteo, Larry A.; Poncibo, Cristina and Cannarsa, Michel. Cambridge University Press, 2022
https://doi.org/10.1017/9781009072168

B-Cube-AI, "Cutting-Edge AI Crypto Trading Bots", Accessed 2023 https://b-cube.ai/

Buckner, Cameron and Garson, James. "Connectionism: The Stanford Encyclopedia of Philosophy" Last modified 2019, https://plato.stanford.edu/archives/fall2019/entries/connectionism/

Campaigns of the World, *Campaigns of the World: Huawei Facing Emotions* (Accessed 2023) https://campaignsoftheworld.com/technology/huawei-facing-emotions/

Coeckelbergh, Mark. "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability." *Science and Engineering Ethics* Vol. 26 (2), (2020) https://doi.org/10.1007/s11948-019-00146-8

Crown Prosecution Service, *Legal Guidance, Violent Crime: Corporate Manslaughter* (July 2018)
https://www.cps.gov.uk/legal-guidance/corporate-manslaughter

DeepMind Technologies, "Playing Atari with Deep Reinforcement Learning.", Last modified 2015
https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf.

European Commission, *Questions & Answers: AI Liability Directive,* Last modified September 2022
https://ec.europa.eu/commission/presscorner/detail/en/QANDA_22_5793

European Commission, "Regulatory Framework Proposal on Artificial Intelligence", 2021 https://commission.europa.eu/document/f9ac0daf-baa3-4371-a760-810414ce4823_en

Expert.AI, "Expert.AI: What is Machine Learning? A Definition", Last modified March 2022

https://www.expert.ai/blog/machine-learning-definition/

Fenwick, M. and Wrbka, S. "AI and Legal Personhood." In *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, eds. Larry A. Di Matteo, Cristina Poncibo and Michel Cannarsa. Cambridge University Press, 2022) https://doi.org/10.1017/9781009072168

Kurki, Visa A.J. *A Theory of Legal Personhood*, Oxford Academic, 2019

https://doi.org/10.1093/oso/9780198844037.001.0001

Law, Jonathan and Martin, Elizabeth A. *A Dictionary of Law (7th Edition), Online Version*, Oxford, 2014

MacCormick, Neil. *Institutions of Law: An Essay in Legal Theory*, Oxford University Press, 2007

Novelli, Claudio. "Legal Personhood for the Integration of AI Systems in the Social Context: A Study Hypothesis." in *AI & Society,* Vol 1, (2022) https://doi.org/10.1007/s00146-021-01384-w

Searle, John R. *The Construction of Social Reality,* Penguin, 1995

Wittgenstein, Ludwig. *Philosophical Investigations*, eds. and trans. Hacker, P.M.S. and Schulte, Joachim, Oxford: Wiley-Blackwell. 1953, ed. 2009

The World Bee Project, *The World Bee Project: Protecting Pollinators, People & The Planet* (Accessed 2023) https://worldbeeproject.org/

Cases

Citizens United v. Federal Election Commission [2010] *558 U.S. 310*

R v P&O Ferries [1991] 93 CAR 72
Legislation

Corporate Homicide Act (2007) (CMCHAct)
https://www.legislation.gov.uk/ukpga/2007/19/contents

European Commission, *Council Directive 85/374/EEC of July 25 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products (Product Liability Directive)* OJ L210/29 (1985)

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31985L0374

European Commission, *Proposal for a Directive of the European Parliament and of the Council 2022/0303 of 28 September 2022 on adapting non-contractual civil liability rule to artificial intelligence (AI Liability Directive)* OJ C496, (2022)

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496

European Commission, *Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts* 2021/0106 (COD) (2021)
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

National Conference of Commissioners on Uniform State Laws, "Uniform Electronic Transactions Act (UETA)" (1999) 1/20/00

http://euro.ecom.cmu.edu/program/law/08-732/Transactions/ueta.pdf