



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

## ARCHIVIO ISTITUZIONALE DELLA RICERCA

### Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Legal personhood for the integration of AI systems in the social context: a study hypothesis

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Legal personhood for the integration of AI systems in the social context: a study hypothesis / Claudio Novelli. - In: AI & SOCIETY. - ISSN 1435-5655. - ELETTRONICO. - 37:(2022), pp. 1-13. [10.1007/s00146-021-01384-w]

*Availability:*

This version is available at: <https://hdl.handle.net/11585/844336> since: 2022-01-08

*Published:*

DOI: <http://doi.org/10.1007/s00146-021-01384-w>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

**Novelli, C. Legal personhood for the integration of AI systems in the social context: a study hypothesis. *AI & Society* (2022).**

The final published version is available online at: <https://doi.org/10.1007/s00146-021-01384-w>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

*This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)*

***When citing, please refer to the published version.***

# Legal personhood for the integration of AI systems in the social context: a study hypothesis

**Author information.** Claudio Novelli, PhD candidate in Law, Science and Technology, CIRSFID – Alma AI, University of Bologna, Bologna, Italy; and PhD candidate in Droit, Faculty of Law, Economics and Finance (FDEF), University of Luxembourg, Luxembourg City, Luxembourg. [claudio.novelli@unibo.it](mailto:claudio.novelli@unibo.it); ORCID: <https://orcid.org/0000-0003-4193-8604>

**Abstract.** In this paper, I shall set out the pros and cons of assigning legal personhood on artificial intelligence systems (AIs) under civil law. More specifically, I will provide arguments supporting a functionalist justification for conferring personhood on AIs, and I will try to identify what content this legal status might have from a regulatory perspective. Being a person in law implies the entitlement to one or more legal positions. I will mainly focus on liability as it is one of the main grounds for the attribution of legal personhood, like for collective legal entities. A better distribution of responsibilities resulting from unpredictably illegal and/or harmful behaviour may be one of the main reasons to justify the attribution of personhood also for AI systems. This means an efficient allocation of the risks and social costs associated with the use of AIs, ensuring the protection of victims, incentives for production, and technological innovation. However, the paper also considers other legal positions triggered by personhood in addition to responsibility: specific competencies and powers such as, for example, financial autonomy, the ability to hold property, make contracts, sue (and be sued).

**Keywords.** personhood; responsibility; AI; civil law; social costs; insurance.

## 1. Introduction

The problem of embedding artificial intelligence systems (AIs) in society is addressed here from the legal perspective; namely, as an issue of managing, through legal instruments, the social risks and costs posed by systems that, although more precise than human beings, are not infallible.

The user of a self-driving vehicle, a consumer making a contract with a bot, a physician supported by a diagnostic system; each of these interacts with artificial entities that apparently play an active role. The behaviour of such entities (AIs)

seems to be autonomous, meaning that it may not be the result of instructions predetermined by the human designer as a specific reaction to a given circumstance. Rather, it is the result of these systems' abilities to autonomously process information, also by virtue of self-learning algorithms, and to then provide appraisals and decisions. Also, each artificial system can be part of a dense network of agents, in which information can be shared and operational decisions coordinated (think of a fleet of autonomous cars).

A number of questions concerning legal liability are then raised. From which actors is it appropriate to expect explanations and justifications for the behaviour of such a system, and which of them should bear the consequences should such behaviour prove harmful? There are several candidates: e.g. the owner, the user, the programmer, the management of the development team, the testing staff. In some cases, however, responsibility cannot be traced back to any particular person; this essay focuses on those situations where a failure of an AI system depends largely on its operational autonomy.

Generally speaking, to treat AIs as mere cognitive tools implies adopting conventional liability schemes. Some of these schemes succeed in dealing with situations where the role and potential blame of human parties can be traced, while they risk being unsuitable for those situations where the damage is caused by an incident of the artificial agent's autonomous agency and the distributed knowledge it is driven by.

Treating AIs as persons under the law would mean making them entitled to a series of legal positions (or just to a single one according to a *thin* conception of legal personhood) (Pietrzykowski 2017). This may imply, among other things, assigning direct liability to the technological artefact, however bizarre this may at first appear. Indeed, the hypothesis of an 'electronic personhood' for AIs, also presented in a European Parliament resolution in 2017, has not yet found political consensus nor has it been endorsed by the European Commission. The rationale is twofold: (a) there is a widespread view that some of the risks associated with AI can be addressed with existing liability rules, or with preventive measures and controls (as in the recent EU proposal for an AI Regulation); (b) the EU lacks the competence to decide what counts as a 'person' within national legal systems since this decision is up to the Member States. Yet, this does not prevent us from exploring the regulatory hypothesis since traditional legal schemes raise some relevant concerns, as will be seen (section 3), if not for the current generation of AIs then for future ones.

I shall mainly consider a functional approach to legal personhood: among the reasons that may justify conferring personhood to an entity, there is the expected utility of the better distribution of risks, benefits, and externalities as realised by virtue of allocating legal positions to a separate point of imputation. Similarly to what happens for collective legal persons, I claim that, under certain circumstances, the allocation of legal positions (especially liabilities) directly to an AIs is desirable when, as compared to conventional legal schemes, it ameliorates the managing of

social costs and opportunities arising from these technologies – as well as providing the right incentives. .

I shall then speculate on the content of a potential legal personhood status for AIs, demonstrating both the pros and cons of this regulatory option.

## **2. What is an Artificial Intelligence System?**

There are many official definitions of what artificial intelligence systems are, provided both by independent organisations and by national legislators. I would take as a reference that provided by the expert group appointed by the European Commission in 2019 (AI HLEG): “Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions [...]”.

Thus, when talking about AIs one generally refers to software running either in a simple computer – or in a computer network – or in a physical architecture – i.e. the hardware component – made up of sensors, computer vision devices, etc. The former might be defined as generic ‘AI algorithms’ or ‘software agents’ while the latter might be called ‘robot agents’. Beyond software or hardware, the core element of both these technologies is *data* on which AI systems are built and trained.

Concerning robot agents, there are several industrial applications, both for logistics and manufacturing. There are also interesting uses of robots for educational and recreational purposes, as well as to assist vulnerable people, and to provide surgical assistance for high-precision operations (Ben-Ari and Mondada 2018).

For the purpose of this essay, only a few of these robot agents would end up in the set of AIs potentially eligible for legal personhood. A robotic arm mounting a tractor, however autonomous, is unlikely to achieve the complexity for which it would make sense to consider having a legal status other than that of a mere ‘thing’ unable to pursue goals autonomously, without a higher degree of operational flexibility, social impact and risk tolerance. On the other hand, physical robots and drones that provide, for instance, transport or assistance services might.

Looking at ‘non-robotic’ AI systems, the applications are even more numerous. This group includes both expert systems, such as computer programmes with specialised expertise in given areas, such as medical diagnostic systems and more flexible technologies like *machine learning* (ML) algorithms; the latter are

based on example-driven learning methods that allow programs to improve their performance autonomously with experience (Mitchell 1997).

Among the software agents built on ML algorithms, particularly relevant for the legal discipline are trading and contract agents: electronic agents to whom human users delegate cognitive skills instrumental to activities of contract formation, negotiation and conclusion. Often, such software agents are capable of determining much of a contract's content without human supervision until the final agreement is reached (Andrade et al. 2007; Chopra and White 2010; Dahiyat 2021). Precisely with self-learning capacity, intelligent agents of this kind employed in dynamic environments can infer standards of conduct that do not comply with the user's instructions. Some examples are advanced digital bots used in online commerce, or financial bots, insofar as they process information and make choices between epistemic and practical alternatives, make consequential changes in the digital or physical world, and engage in communication.

AI technologies we are concerned with can deliberate according to evaluations that are, to some extent, independent of those of the designer. In a sense, AIs have a degree of epistemic and practical authority over their own behaviour. These features also enable AIs to cope with stochastic environments, adjust to the scarcity of information, obtaining new information from the environment and interactions with other agents. AIs can integrate their initial epistemic and practical knowledge when this proves inadequate for the context in which they operate, and thus act on the basis of an expanded knowledge not necessarily accessible to their users and developers.

In contrast to other artefacts, by virtue of the described abilities, AIs can be assigned tasks without ongoing human oversight. The other side of the coin, however, is the lack of control by human users, crucial if AIs are involved in legally relevant activities. Often, it is almost impossible to predict which of the many courses of action the artificial agent will pick in the presence of changing circumstances..

It seems difficult to provide an identikit of AI systems that are potential candidates for acquiring legal personhood based solely on technical peculiarities. My proposal is to embrace a multifactorial approach since what makes the research question meaningful is the occurrence of a set of socio-technical factors: i.e. along with technical skills related to knowledge extraction, self-learning, and self-adaptation, the level of risk tolerance of the AIs and the social infrastructure in which the same operate – and on which decisions impact – are also key elements.

AIs that meet these requirements already exist, as is the case of many trading agents using machine or reinforcement learning algorithms. Imagine, for example, that such an agent trades in used cars and self-induces an association between optimisation of its utility function and some unfair commercial practice, like 'bait advertising'. The legal consequences flowing from such unlawful conduct are those that need to be addressed.

## 2.1. The legal salience of AIs

This raises questions about the legal imputation of AIs' actions as well as a more general question on the juridical classification of such systems as mere 'things', rather than as persons. Especially if one considers that AIs can be treated as true intentional agents – i.e. by virtue of the intentional stance (Dennett 1987) – so as to explain and predict their behaviour through the assignment of cognitive states: it is possible to assume that they act on the basis of information about one or more states of the world (epistemic states), to achieve certain goals (conative states) through a rational and flexible decision process (decision states).

Under some circumstances, the conduct of AIs can only be interpreted as the output of intentional states independent from those of the human beings involved. This has consequences that legislators should address in terms of how third parties interact with AIs, e.g. these parties must rely on the intentional attitudes directly exhibited by the system, irrespective of the upstream intentions of human controllers.

This broad theme is addressed by this essay with particular attention to the subjective legal position of civil law liability, in the context of the hypothesis of attributing legal personhood directly to AIs. Indeed, in the current legal framework (European and beyond), even if AIs act with a certain degree of autonomy, they do not bear the legal consequences – rewards or responsibilities – of doing so. Such a system does not seem optimal, at least for cases of unpredictable failures, and the question arises as to whether it is appropriate to assign responsibility to AIs and, even more, how to do so.

## 3. The limits of conventional liability schemes

As long as we consider an AIs as a mere cognitive tool, civil law liability for damages caused by 'things' usually falls on the shoulders of the custodian – either the owner or another controller – or the designers-producers. Generally, we have at hand both fault-based and non-fault-based (or even no-fault at all) liability patterns. These patterns differ in the weight they attach to psychological commitment and in the exonerating evidence, and both show some problems if applied to AIs.

Fault-based schemes, which are mostly negligence liabilities, usually hold liable the person who has failed to observe the appropriate *standard of care*. This type of approach is likely to work where the duty of care, i.e. the conduct required, is discoverable and virtually ascertainable in court. Judicial definition of duty of care, when it comes to new technologies, is a key passage, but one that requires adequate knowledge of the technology involved. Also, judicial assessment of the standard of care helps to calculate the inherent risk and the degree of acceptance of a given technology (i.e. risk knowledge) (Zech 2021). Where the failure is traceable to the accountable party – be it the programmer, the manufacturer, the data

developer or the operator – the negligence-based model incentivises diligence, including in third parties.

However, effectively setting the standard of care tends to be compromised by the operational paradigms and knowledge extraction methods of modern AIs. As already mentioned, the capacity for autonomous deliberation and action is matched by a lack of control of the human user. This is particularly evident for more sophisticated AIs that also exhibit a kind of ‘creativity’ in the decision-making process. One of the most famous is AlphaGo, the software that has learned to play the game of Go in a professional manner (Silver et al. 2016). Although AlphaGo's programmers knew how the basic neural network worked and also knew the data on which the system had been trained, at several points they could not predict the exact move or game tactic that the software would implement.

Therefore, this scheme is not suitable for all those cases in which harmful consequences are caused by an AI system that has been used appropriately by users and, incidentally, was not defective at the time of its introduction on the market. These are situations in which an AIs may have engaged in harmful or unlawful conduct despite meeting the requirements of the state of the art and being properly used.

It seems then that two of the cornerstones of personal responsibility are missing: a sufficient degree of awareness of what happens (the *epistemic condition*) and control over the actions performed (the *control condition*) (Coeckelbergh 2020). These conditions are poorly met both by users and by the developers themselves, who are often unable to predict the specific course of action an AI will take.

If a standard of care (and development risks) cannot really be relied upon or defined due to information gaps, the profitability of negligence liability decreases. A liability dilemma then arises: either we have human beings liable for actions that they may not have intended to take, nor could reasonably have prevented, or a liability gap occurs, as no one can really be held liable. Both scenarios, unsurprisingly, increase the regulatory uncertainty that can affect the production of AIs – hence technological improvement – as well as access to justice for the injured party (Zech 2021).

The other group of liability law solutions considered above, go in the direction of providing legal certainty in the first place, i.e. strict liabilities like those for products or ultra-hazardous activities. The party held strictly liable, aware of the embedded risk, may internalise the costs to be incurred or pass them on to others (e.g. final consumers) and consequently plan their businesses with greater certainty. The burden of proof would be significantly eased for the victim, who would have a better chance of being awarded compensation. In addition, the party exposed to liability will be incentivised to work on improving technology and enhancing safety and reliability standards, thus ultimately benefiting the general interest.

Although this approach seems suitable for less widely used high-risk AIs, it may be unduly burdensome to apply it to all AI devices, i.e. to systematically

offload the costs of accidents onto a single party. If the latter is the manufacturer (or the coder), such mechanism might disincentivise production and discourage consumer diligence as someone else will always be held liable; while if the unique liable party is the end user, strict liability might act as a brake on purchase and obviously would not provide incentive for manufacturers to improve their products.

As a matter of fact, tort law rules play a role in the size of the industry and the price of goods: negligence liability systems are associated with lower costs in supplying goods, since the producer can escape liability by meeting legal standards of care, whereas strict liability systems are associated with higher costs. As a consequence, industry will be allowed to grow with negligence-based liability, because of the effect on the demand curve of a lower price of goods (Cooter 1991). A downturn for a relatively young industry such as AI could have detrimental effects including, for instance, a lack of social trust in new technology and a loss of innovation.

On top of that, relative strict liability schemes still require proof of causation, which would not be always straightforward for AIs “[...] due to the increasing connectivity, situations may arise where distant contributors (e.g. by supplying faulty data) cannot be detected. In cases where an immediate causal person acting in accordance with duties of care can be determined (e.g. the operator of a damaging hardware component), while a more distant contributor acting in breach of duty is not held liable, strict liability may result in an incentive for lower levels of care for the distant parties” (Zech 2021, 7). Therefore, the typical effect of the strict liability scheme, i.e. the internalisation of the costs (of accidents), is undermined as it can be too onerous for the victim to identify the perpetrator(s) and the causal link in order to file a lawsuit. If the victim is disincentivised to claim compensation, while perpetrators benefit from areas of impunity, then tort liability rules do not function properly.

To sum up, in borderline cases, where harmful conduct emerges as an expression of AIs' operational autonomy, custodians cannot be blamed for negligent supervision, and strict liability risks being unfair, or at least too severe, especially where the damage that needs to be redressed is unexpectedly high. In particular, strict liability placed on the shoulders of producers or programmers risks discouraging production and innovation. Thus, conventional liability schemes risk being unable to introduce incentives that both guarantee continue innovation as well as victims' rights to compensation.

One way of securing creditors through separate assets could be provided by conferring a separate legal personhood on AIs. This could also allow the AI agent to hold rights and obligations and to enter into contracts on its own behalf and to produce valid legal effects on third parties, possibly with complete financial autonomy. The problem is whether conferring autonomous legal personhood on AIs is an efficient regulatory strategy for dealing with opaque and unpredictable situations.

#### 4. The functions of legal personhood on AIs

The status of legal personhood links a centre of interests to bundles of rights and duties and is triggered by multiple and divergent incidents. It is frequently claimed that legal personhood can be conceived as a cluster concept, as we will elaborate in the second part of this essay. The difference between the status of *things* and that of *persons* – common to most legal systems – is anchored in the belief that persons cannot be objects of use, have their own interests, exhibit agency and their conduct, with the resulting effects, can be attributed to them; on the contrary things are objects of use: they do not possess will, interests, competence or accountability (nor then legal standing) (Benson 2002). To some extent, AIs challenge this distinction since they are equipped with an epistemic and practical authority over their behaviour.

However, more than a property-based approach I would subscribe a functionalist one, i.e. personhood status is mainly an instrument for allocating legal positions. Contrary to a meaning of personhood that ends up being co-extensive with the concept of ‘human being’ or with that of ‘intentional subject’, from a functional-legalistic perspective a person is only a point of imputation of legal positions. Then the question to be asked is: are there any practical advantages to creating a detached point of imputation of legal positions to serve an artificial agent? This approach has characterised doctrinal reflection on the conferral of legal personhood on corporations too, which is mainly justified by the minimisation of transaction costs, efficient coordination of several individuals and the opportunity to differentiate liability, risk and tax treatment (Freund 2000). Such legal personhood also simplifies underlying relationships, both between members and with external parties (e.g. creditors), through an appropriate legal addressee rather than through a network of contracts. Finally, one thesis held in doctrine is that the legal personhood of corporations, and thus the partitioning of assets to secure both members and creditors, has historically produced stimuli for economic growth (Hansmann et al. 2003).

I believe that similar functional gains can be realised through the legal personhood of the AIs: legal simplification, transparency, efficient distribution of externalities produced by AIs, better protection of injured parties and the benefit to society of having such artificial agents being more self-sufficient in activity. The side effect is to stimulate the growth of such an innovative and profitable industry.

Although some of the externalities produced by an AIs can be addressed through the rules of tort law without necessarily depriving AIs of the status of ‘things’, traditional solutions do not lead to efficient results for other, more problematic, situations: e.g. unpredictable accidents. The comparative advantage of treating AIs as persons can then be to hold them directly liable. Let us see how.

Sanctions for civil wrongdoings are typically pecuniary and accordingly the status of personhood under civil law can be sensibly claimed only if there is an asset assigned to the artificial agent. Thus, the hypothesis to be contemplated assumes

that an AI has its own assets over which victims and creditors will be able to exercise their right to compensation in the event of wrongdoing. Having an artificial agent's own assets would allow the limitation of liability and asset segregation – hence risk mitigation – that frequently informs the organisation of economic and collective activities in the form of juridical persons. In this way, depending on how the status of legal personhood is designed, it will be possible to spread internalisation costs over several stakeholders while segregating their individual assets, with the associated rewards in terms of economic incentives for production and innovation.

Moreover, as far as the technical-legal aspect is concerned, having a single and univocal point of attribution of legal positions is both a means of simplification and a guarantee for the victims (whether third parties or the users themselves). Indeed, legal personhood – and so their legal standing – would make it less costly to identify the accountable party in view of the litigation, compared to the requirements of negligence or strict liability rules on causation and standard of care (particularly due to the so-called ‘many hands’ problem). Much less difficulties would be encountered in proving the causal link because the ‘owner’ of the action – whether intentional or not – would be the AI in the first place. Instead, the benchmark for assessing diligence would depend more on the technological state of the art, perhaps declared through certification standards, rather than on the conduct of the human party. From this perspective, it would make sense to combine compulsory insurance mechanisms where the state of the art does not help to protect the victim appropriately. Finally, if the rules on corporate insolvency were followed, separate assets could ensure that the victims of AIs would have priority over the personal creditors of shareholders.

To benefit from the status of legal personhood, AIs should then be entered in a dedicated register, which presupposes that they meet the requirements of a certification and registration procedure (Sartor 2002). While registration might identify the ‘domicile’ of the system and disclose the amount of the patrimonial warranty, certification will report its technical features, e.g. degree of autonomy, self-learning capacity, error rate, sources and methods of knowledge extraction and so forth. These procedures would improve the transparency and reliability of the AIs introduced on the market.

At this stage, it makes sense to ask whether asset independence serving the capacity of an AI agent to act has advantages or whether, on the contrary, it should be viewed with concern. Clearly, we cannot make an absolute statement here, and much will ultimately depend on the rate of accuracy and success of AI technologies, as applied to a specific field. While for some systems it may be possible to take advantage of operational self-sufficiency – with the associated ability to enter into legally relevant interactions (e.g. high-frequency trading agents, but also some self-driving vehicles), for others it may remain preferable to tie behaviour to the supervision of a human being, e.g. autonomous weapons.

## 5. The AIs' legal personhood

The resolution of the European Commission with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) is well known in the relevant literature for suggesting that the most advanced AI systems should be given an 'electronic personhood'.

To take up the European Parliament's advice, one should have an approximate idea of the type of personhood status being tested. Personhood does not necessarily have to assign a fixed set of legal positions, e.g. rights, duties, powers, immunities, liabilities, or liberties (Hofheld 1913). The set of subjective positions may vary according to the entity considered, and thus we will have cases in which the personhood attributes only rights – e.g. natural and animal entities (the latter under some legal systems) – or, rarely, only obligations – compensatory funds.

Before even designing the specific 'dress' of civil law personhood of the AIs, we need to clear up some ambiguities: one can speak of legal personhood both in *functional* and in *constitutional* terms.

So-called constitutional personhood denotes that legal status that guarantees the protection of individual rights and fundamental freedoms of human beings; rights and freedoms that are precisely recognised by the constitutions of legal systems. These are mainly, but not exclusively, prerogatives associated with being human, e.g. the right to bodily integrity, freedom of thought, freedom of religious belief, freedom of speech, etc. If AIs gradually acquire human attributes, then they could progressively acquire the legal (and moral) status of human beings; but it would probably make little sense to imagine that current AIs would enjoy such personhood status, both for factors related to cognitive competence and for those related to the degree of social integration of such artefacts (Andreotta 2021).

For the purposes of this paper, a functionalist approach to legal personhood is preferred, and therefore I will appeal mainly to instrumentalist/functionalist justifications and less to essentialists/constitutional ones. Although a break between these dimensions is not necessarily desirable. As Lawrence Solum observes, the conferral of many constitutional rights and freedoms, for instance, on AIs can be justified on a functional basis: “Granting AIs freedom of speech might have the best consequences for humans, because this action would promote the production of useful information. But assuming a different justification for the freedom of speech can make the issue more complex. If we assume that the justification for freedom of speech is to protect the autonomy of speakers, for example, then we must answer the question whether AIs can be autonomous” (Solum 1992).

It is certainly possible to explore some ontological grounds (Solum 1992; Chopra and White 2011), but in the perspective of regulatory policy it is not a necessary stage for at least two reasons: (1) there is nothing to prevent us from assuming that personhood is a mere legal instrument, and can therefore be invoked to serve objectives that are independent of AIs as such, like cost allocation and the pursuit of social opportunities; (2) it is not always the case that legal personhood is

grounded solely on distinctively human attitudes: attributes like conscience, empathy, feeling or moral sense may not be necessary (nor sufficient) conditions for personhood (Matthias 2008).

### 5.1. The content of AIs' legal personhood

What specific bundles of rights and duties could make up the (functional) personhood status of an AIs? In some cases, personhood consists of the entitlement to a series of legal positions that must be enforced through or on the initiative of other legal persons. Some classic examples are those of corporations, children, unborn children or persons with impaired capacity, who find in their principal (representatives, parents or guardians) those subjects protecting their interests and implementing their personhood prerogatives (e.g. suing or being sued). Animals, natural entities or idols might be added to the list in the future (Brunet 2019; Pietrzykowski 2018).<sup>1</sup> Other legal persons, on the other hand, deliberately dispose of their own bundle of legal positions: e.g. adults of sound mind.

These two kinds of status are also known as *dependent* and *independent* legal personhood: "A dependent legal person can only act through the agency of another legal person in exercising some or all of its legal rights. An independent legal person is not subject to any such restriction and is said to be *sui juris*" (Chopra and White 2011). The status of the independent legal person generally coincides with constitutionally guaranteed protection. As anticipated, this status is more entangled with ontological traits of the entities under consideration and, when applied to AIs, ends up being more of a thought experiment than a proper regulatory enquiry.

The scenario of an AIs dependent legal personhood is theoretically more realistic, as it would definitely not be the first time that statuses of this kind have been attributed to artefacts (e.g. corporations, ships, idols, patrimonies, etc.), and seems quite in continuity with the idea of holding AIs as parts of collective legal entities. Such a status might consist of passive legal positions alone: the right to bodily integrity, to not be the object of someone else's property, to own property, to receive by inheritance or donation, to participate in succession, to be entitled to compensation for damages, to be represented in a legal transaction or in court, and so forth. These legal stances constitute what is generally known as *capacity for rights* or *passive legal capacity*: "Passive capacity is properly identified as an entity's capability in law to be the beneficiary of some legal provision or provisions, in the sense that these provisions are interpreted as aiming at protecting such an entity from some harm or at advancing some interest or another of that entity" (MacCormick 2007). Those who hold this type of legal personhood are generally only *right-bearers*.

---

<sup>1</sup> There are legal systems in which such statuses have been already recognised, to the extent that the term environmental and animal personhood is used in this regard. Some examples are New Zealand and India, which have recognised legal personhood for rivers.

Nevertheless, dependent personhood can also consist of *active* legal positions; for example, although corporations can only act through representatives, they have a series of active powers, i.e. they have the *capacity to act*: the power to produce (valid) legal effects both in one's own legal sphere and in that of others and, eventually, to incur the corresponding liabilities. Holders of this type of personhood are not only right-bearers but also *duty-bearers*.

The set of rights and duties that make up the capacity to act is very broad, also changing according to the specific entity and can be split into two sub-groups, *transitional capacity* and *liability capacity* (MacCormick 2007).

Transitional capacity describes the group of powers to enter into or create legally salient relationships through one's own action and deliberation, producing valid legal effects able to bind oneself and others. The network of relationships and legal effects may be manifold: contracting, transferring property, suing, donating, making a valid will, contracting marriage, voting, serving as a trustee, registering domicile, and so on; not to mention all the prerogatives of collective legal persons (e.g. merging). Clearly, it may not make much sense to attribute some of these positions to an AIs, as is the case with voting rights. It might, though, make sense that an AIs would not only be the assignee of an asset but would have the ability to dispose of that money, more or less freely, for transactions and investments. Transitional capacity makes it possible to participate in such interactions by producing valid legal effects, by creating positions of claim for judicial enforcement in cases of violation or non-fulfilment (MacCormick 2007).

Liability capacity, on the other hand, is the susceptibility to legal imputation for civil wrongdoings (or criminal offences). It makes sense to distinguish it from mere capacity to act since sometimes subjects capable of producing valid legal effects may not be held liable for any unlawful consequences. One example is that of minors, including those under the age of 14, who are frequently authorised to take legal action for small purchases or to freely dispose of the assets at their disposal even though they do not enjoy full capacity to act..

Given that dependent personhood status can be so articulated, it may not be necessary to confer *sui juris* status on AIs, i.e. a complete package of legal positions, to guarantee these systems a certain degree of autonomy and to handle the thorny issues of liability. Indeed, it might be sufficient to ensure that the AIs have assets that are separate from those involved, just as limited liability companies remain legal entities dependent on their members.

How far this asset will be able to ensure the optimal allocation of resources, given the different positions at stake, will depend on the contributions that flow into it and on how the legal personhood of the machine itself is designed. One solution could be to bring into the AIs assets financial contributions from all human parties involved in development, service provision, production, and use of such products; almost in the form of a compulsory insurance or compensation fund, thus reproducing a no-fault system (i.e. absolute strict liability) (Ziemianin 2021).

In contrast to simple insurances or funds, however, the AIs could also actively dispose of the assets conferred on it, and possibly generate new gains with which to increase the security of any creditors or even just amortise management costs, e.g. lowering participation fees. Revenues may come either from the AI agent's own activity and services – think of commercial and financial agents, but also diagnosis systems and creative agents (whether of software or works of art) – or through investments complementary to the main employment. A share of these gains could be reserved for the stakeholders and another directly for the AIs, with the option of reinvesting them or simply keeping them as collateral. Wealth accumulated by AIs could then also be taxed (Oberson 2019; Huettinger and Boyd 2019; Atkinson 2019).

Much of the actual financial autonomy will depend on the powers granted and the degree of independence of the resulting legal personhood: i.e. in the case of a dependent personhood, it is to be expected that earnings will be put primarily at the service of the stakeholders who will set the guidelines for management and financial prudence. Alternatively, it could be the case that, after raising additional funds, AI agents themselves buy their own insurance: “If the AI could insure, at a reasonable cost, against the risk that it would be found liable for breaching the duty to exercise reasonable care, then functionally the AI would be able to assume both the duty and the corresponding liability” (Solum 1992, 1245). If AIs were to take out their own insurance, this would perhaps allow them to exploit the benefits of both systems: a probably higher compensation fund with which to secure victims, and the segregation of the personal assets of the individuals involved.

This personhood-insurance mechanism would bring benefits similar to those derived from the creation of juridical persons, e.g. asset segregation, transparency, judicial simplification and creditors' pre-emption. The system will undergo registration, certification procedures and could use a digital signature tool to authenticate itself.

The first procedure will serve to ensure reliance on assets, both counterparties interacting with an AIs – e.g. for commercial purposes – and active contributors will be made aware of the amount of money available to the artificial agent. On the other hand, certification, and thus the declaration of the technical peculiarities about system operation, risk and autonomy, may help to set the insurance premium charged to the parties concerned. Finally, a digital signature could be used to identify the individual AIs, to contrast some tracking issues – risk of software duplication – and ultimately make the legal interaction valid.

What is more, victims would no longer have to prove either the causal link or negligence of the human actor, as it may suffice to prove that the damage was caused by a source covered by this sort of insurance (Solum 1992). Therefore, it will be decisive to bring clarity to what events are covered by the AIs' asset guarantee. It is accidents due to the AIs' autonomous agency, opaque knowledge processing and the extraction of information from remote and untraceable sources that pose intriguing liability issues. It has to be said, to be fair, that this could appear

just as complicated. Indeed in the model I am advocating, which does not exclude liability of ‘humans in the chain’ for negligence and malpractice, it would still be necessary to show that the origin of the failure is wholly attributable to the machine; and this may not be an easy task.

Anyway, if these are the main events covered by an AI’s personal asset/insurance, then this requires two things on the liability side. The first is that such events should not be excessively infrequent, otherwise the economic justification for the efficient distribution of costs through a shared stake in the assets of AI’s risks becoming untenable. If such events were indeed rare, then an ordinary non-fault-based liability scheme (even absolute) might suffice. But this point would require further reflection. The second requirement is that if only such events are covered, then it would be appropriate for such a system to be combined with fault-based liability mechanisms; patchworks of this kind are constantly being tried out by different legal systems (Wagner 2012).

A mixed discipline must then discriminate along two axes: both (1) the AI’s model, e.g. only those showing sophisticated processing and learning capabilities will be entitled to legal personhood and (2) the outcome of an AI system’s conduct – e.g. if the outcome is caused by a clear mistake in programming, manufacturing or misuse of the system it will be desirable to apply traditional liability methods (e.g. product liability or vicarious liability), so as not to encourage the concealment of malpractice or foster suboptimal equilibria of innovation and professional diligence; otherwise the direct personhood/responsibility of the AI’s would remain.

As far as the first axis is concerned, a three-stage approach can be adopted: ranging from the minimum level of autonomy for which the legal status is essentially that of a cognitive tool (to be associated with custodian liability maybe); the intermediate level for which some legal autonomy begins to be conferred (maybe housing the AI’s in human-controlled juridical persons); and the maximum level which is associated with the proper status of individual legal personhood (Dahiyat 2021).

A further criterion to be used in conjunction with mere operational autonomy could be the risk of social impact (and harm) of the AI agent: so that, for instance, robots with a high degree of autonomy but a negligible capacity to cause considerable social damages would still be treated as cognitive tools. In this latter case, conventional liability rules would continue to apply.<sup>2</sup> A risk-based regulatory framework recently proposed by the European Commission is worth mentioning in this spirit. AI’s are classified on three levels of risk: (r1) unacceptable; (r2) high; (r3) minimal. Unacceptable risk is associated with: “All AI systems considered a clear threat to the safety, livelihoods and rights of people will be banned, from social scoring by governments to toys using voice assistance that encourages dangerous

---

<sup>2</sup> The risk criterion and the adoption of a similar system, was proposed in the expert group report on AI liability to the European Commission: Liability for artificial intelligence and other emerging digital technologies (2019): <https://op.europa.eu/it/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1>

behaviour”.<sup>3</sup> Systems that, for example, employ subliminal techniques to subconsciously guide a person's behaviour are therefore banned.

High-risk AIs include both those “[...] intended to be used as safety component of products that are subject to third party ex-ante conformity assessment” and “other stand-alone AI systems with mainly fundamental rights implications [...]”.<sup>4</sup> Some applications of these systems are, e.g., biometric identification systems, critical infrastructures, credit scoring for obtaining a loan and others.

According to the regulation, manufacturers of these high-risk AI systems will have to fulfil a number of obligations to place them on the market. In particular, they will have to provide adequate risk assessment and mitigation systems; ensure that datasets are constructed to minimise risk and possible discriminatory results; and guarantee the traceability of results, including by providing all necessary information to competent authorities to assess compliance. In addition, the same manufacturers should take care to provide clear and adequate information to the user, including on risk-minimising supervision.

Finally, at the minimal risk level are those systems that do not affect, or at least do not risk compromising, fundamental rights and values, which are also the most common on the market today: predictive maintenance systems, spam filters and AI for video games. The second axis, i.e. the distinction based on the type of AIs’ failure, is likely to be more complex because it is often unclear which conduct actually emerged from artificial agency and which from human mistake or recklessness. However, negligent behaviour can be attributed to either the manufacturers, programmers or users – or to a combination of them – because even if the latter have much less supervisory powers than the former, it cannot be ruled out that a malfunction is due to a lack of monitoring or diligent use on their part. Where this reconstruction is not possible, direct liability of the AIs might be preferred.

A further idea to bear in mind, which can be seen as an alternative to the mechanism envisaged so far, limits the area of assets conferred on the AIs, and therefore of legal personhood, not to liability for any unforeseeable damage, but only for those produced against the owners/users of the systems themselves and not against third parties. The model is that of so-called market enterprise responsibility, i.e. while damage to third parties is borne by the owners or users, damage to the latter is borne by the producers, who place AIs on the market equipped with an asset fund. The result is always the creation of some kind of legal personhood, whereby

---

<sup>3</sup> This is from the European Commission website on ‘Shaping Europe’s digital future’: ‘Regulatory framework proposal on artificial intelligence’, link: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.

<sup>4</sup> European Commission, Brussels, 21.4.2021, Com(2021) 206 Final, 2021/0106(Cod), Proposal For A Regulation Of The European Parliament And Of The Council. Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, p. 13.

the AIs will be directly liable to a limited extent with the assets assigned to them. The advantage is that producers can know the costs of externalities in advance.

The most realistic legal personhood scenario so far remains that of AIs with dependent status, holding mainly duties and responsibilities and fewer rights. It seems premature to discuss conferring independent status on AIs. Samir Chopra and Laurence F. White, in this sense, refer to five attributes: intellectual competence, susceptibility to legal obligations, susceptibility to punishment, contract formation, and economic and property holding capacity. Although many of these skills seem to be within the potential of current AIs, it is worth waiting for technical progress to make more convincing arguments on this point.

To sum up, in line with the approach of the essay on optimal allocation of risks and opportunities, the combination of personalisation and negligence liability rules can serve both prevention of failure due to human negligence and to parcel out the costs of unforeseeable accidents so as not to discourage the production, consumption and technological innovation of AI devices. It is not unnecessary to reiterate that I am not suggesting the indiscriminate appropriateness of the legal personhood regulatory option, but that I am mainly limiting it to the combination of two factors: the unpredictability of the conduct of AIs and the social impact, i.e. those cases in which AIs involve security, fundamental rights and values – situations that the European Commission associates with high-risk AIs. In other circumstances, traditional solutions – producer or user liability in various ways – may be more economically sustainable.

## **6. The case against legal personhood on AIs**

I will briefly review what the main arguments are against the hypothesis of legal personhood on AIs, trying to understand whether they are justified and when they are, whether it is possible to refute them. There are objections of a mainly ontological nature – deferent to an essentialist conception of legal personhood – and objections of an instrumental nature – invoking the uselessness of this choice of legal policy or the inability to achieve the intended objectives.

### **6.1. Ontological objections**

Among the ontological reasons preventing AIs from being conferred legal personhood are generally the cognitive and moral traits of the human condition. The lack of these qualities, according to some authors (Solaiman 2017; Gordon 2020), would prevent AIs from being able to effectively hold the prerogatives and duties associated with legal personhood. It would therefore not only be an inherent problem of what these systems *are*, but of what these systems *do* (Chesterman 2020).

The distinctive properties of the human condition generally assumed to be presupposed by a full and effective right to legal positions are intelligence,

conscience, self-awareness, sensitivity, (the experience of) free will, and moral agency. Doubts, therefore, arise as to whether AIs are intellectually competent enough – or have the right form of competence – to understand legal obligations and to dispose of their legal sphere.

Having personhood, being the addressee of a set of legal norms would then require the capacity for these norms to be understood by the addressee. The lack of proper comprehension is not a problem for juridical persons since it is still the human members who determine how the collective agent will acknowledge the legal rules and behave accordingly; while it could be a major issue for AI agents acting alone.

The first point to make is that the validity of this kind of objection is limited to independent legal personhood, whereas dependent personhood does not appear to be affected. Beyond that, there are some aspects that seem to be underestimated when making statements about the intellectual uniqueness of human beings. First of all, AIs are smart enough to act on the basis of reasons, trying to optimise the achievement of one or more goals, interacting with the external environment and being able to manipulate (also in a creative way) the received information as well as to implement new ones as perceived by the environment and by their own experience. In this sense, it seems possible to describe AIs as intelligent agents, capable of participating intentionally in relevant legal situations – e.g. contracts – without the ongoing supervision of the human principal.

Second, peculiarities of the human condition we still know very little: e.g. the way our conscience works, the role it plays in our self-awareness or deliberations, whether there is a free cause for our actions, etc. We know, of course, that current AIs – and likely next generations as long as this computer architecture is used – are not comparable to humans in terms of consciousness and meaning attribution (i.e. semantics) (Dreyfus 1992). But, as far as legally relevant cognitive capacity is concerned, this is not a conclusive argument against the personhood of AIs.

I would opt again for a pragmatic approach: the problem is not to establish what are the general cognitive peculiarities of human beings, but which of these count for the purposes of legal practice and for the entitlement of active personhood, i.e. composed of both rights and duties. I believe that the key functional attitude here is that agents take legal rules as *reasons for action*: that is, as premises for the practical reasoning which justifies or explains their behaviour (also called reason-responsiveness). AI agents acknowledging such practical and epistemic authority to legal norms would reveal their sensitivity to obligations.

While the protection of self-consciousness, moral maturity, or (the experience of) free will by legal personhood appears as a 'privilege' of some entities only – i.e. adults of sound mind – it would be difficult to make agents who do not perceive legal rules as reasons for action participate in legal interactions. Humans with such deficits are still treated as legal persons since they maintain a moral status, but one may not see any need for it in the case of non-human entities that are not accorded any moral status (yet).

It should be stressed that although artificial agents cannot manipulate semantics, they can still engage in reason-responsive and obligation-sensitive behaviour (Dignum 1999; Castelfranchi et al. 2000; Gelati et al. 2002; Boella and Van der Torre 2007). For this purpose, in fact, it seems sufficient for AI agents to have *instrumental rationality*, i.e. one that is useful for pursuing an interest (one's own or that of the user) by virtue of deliberations that take account of the possible consequences of one's actions and the way in which these consequences – also represented in the form of legal sanctions – may compromise the satisfaction of the interest itself. The fact that the instrumental rationality enabling AIs to recognise and act on legal norms does not coexist with the capacity to be morally involved in them, is not an obstacle to the holding of many of the legal positions embedded in personhood.

Quoting Chopra and White: “Work in deontological logics or logics of obligations suggests the possibility of agent architectures that use as part of their control mechanisms a set of prescribed obligations, with modalities made available to the agent under which some obligations are expressed as necessarily to be followed” (Chopra and White 2011, 166). And AIs may have rational motives, embedded in their basic drives (, to comply with legal obligations, since failure to do so could compromise essential drives as self-preservation or maximisation of expected utility. This last profile also implies some form of susceptibility to punishment.

Some might object that being mechanically compelled to follow rules is not the same as having real sensitivity to obligations, but this is not obvious from a technical point of view: some artificial agents may sometimes be allowed to violate legal norms when the infringement results from their own deliberative process as being more beneficial (Castelfranchi et al. 2000). Such is the case of a software agent that violates privacy rules for fear of catching a virus from a malware. This clearly poses significant problems where the violation is systematically carried out by an AIs for reasons of its own convenience; and the possibility of accidents caused in this way are at the root of some of the liability issues to be solved (also) by the conferral of legal personhood. On the other hand, one might claim that the chances of conforming to virtuous standards of moral conduct are much higher in artificial agents than in natural ones. As some authors ironically suggest, while humans only tend to deontological or utilitarian ideals of moral agency, artificial agents might be totally aligned to those moral standards (Brozek and Janik 2019).

In any case, even if current cognitive qualities of AIs rule out their having proper moral agency – for which the requirement of consciousness and sensitivity might be strongly required (Veliz 2021; Gilbert and Martin 2021) – they seem to meet minimal requirements for participation in legal practice as they can engage in autonomous conduct that can be explained and predicted in terms of intentions and normative beliefs. My point is therefore that from a rigid attitudinal stance AI agents already possess many of the necessary capacities to recognise legal norms, produce legally valid effects and to be held liable for them, but that since they do not have

full moral agency their legal status will not for a long time be comparable to that of natural persons.

For these reasons, and for those related to the functional dimension of legal personhood, the "missing something argument" is not able on its own to defeat the regulatory hypothesis under consideration.

## 6.2. Instrumentalist objections

Instrumental objections mainly point to: the danger of human parties escaping responsibility, the disincentive to improve technology and the dangers of letting AI agents have legal initiative and identification issues (Solaiman 2017; Bryson et al. 2017; Dahiyat 2021). Let us tackle them one at a time.

The phenomenon of evasion of liability is certainly not posed for the first time by AIs, but has long been the subject of debate in the field of company law with respect to the presence of separate legal entities. The limitation of investors' liability that is achieved through corporations gives rise to concerns on the part of creditors when the assets of the corporation are not able to satisfy claims for payment. To counter this abusive practice, the doctrine of the 'piercing of the corporate veil' has taken shape, which makes it possible to overcome the barrier posed by asset segregation and hold the company's stakeholders directly liable (Macey and Mitts 2014). The concern that the same evasive phenomenon also occurs in the case of an AIs' financial inability to repay debts – due to accidents caused or breaches of contract – is heightened by the fact that, in contrast to corporations, in AIs as autonomous legal subjects it would not even always be possible to identify the relevant stakeholders (Bryson et al. 2017).

I am not really convinced by this objection. The attribution of legal personhood has, among other functions, also that of making the underlying relations between the parties involved in the supply (and use) chain of AIs more transparent. Consequently, I do not find any particular obstacle to the application of the 'piercing of the corporate veil' doctrine also to insolvent AIs. Moreover, the patrimonial capacity of an AIs should be in this way mandatorily communicated to third parties – or to be declared at the time of registration on a register of 'electronic persons' – and any operation not adequately covered could be unauthorised (also by design). While all wrongdoing intentionally committed by human actors through AIs would remain covered by conventional forms of liability.

A second type of objection is that assigning legal personhood would not create the right system of incentives and rewards with respect to technological innovation and thus would not promote damage prevention. I believe that this objection is fairly well-founded, since if part of the liability is shared between several parties – including AIs with their own revenues – then it can be expected that the costs of possible accidents will be less on the shoulders of those parties tasked with research and development of safer systems, i.e. manufacturers and designers. In other words, if the cost of making the product safer exceeds the

expected gain from the improvement, as the cost per injury remains stable, then the objective of injury prevention may fail and general welfare suffers.

However, one way to defend against this objection is empirical observation of what happens in other markets. There does not seem to be a vicious cycle like the one just described, for example, in the automotive market simply because: (a) consumers continue to have an interest in buying safer and safer cars and (b) safety devices are required by law to be certified as compliant. The economic interests of the parties involved do not remain invariant as the failure rate of the systems changes: more reliable AIs are more successful on the market, and produce fewer economic losses, both in terms of fees to be paid for damages and in terms of service interruptions, for all parties involved.

Finally, there are criticisms concerning the identification of AIs (as software agents of course) and the dangers of letting them have legal initiative. I believe that both are challenges that can be addressed through technical as well as legal stratagems.

Difficulties in identifying individual AI agents may undermine personhood assignment since these agents have no physical location and may duplicate or reproduce themselves in multiple software. In these cases it becomes impossible to understand which agent is accountable for an action and how it interacts with its environment. In any case, a registration and certification system can curb such problems: “The solution may lie in using digital signatures. Anytime the agent uses a signature, to sign a legally relevant action, it is uniquely identified. That is to say, although software can be copied, keys or signatures are protected in a key-vault, and are protected against copying” (Dahiyat 2021, 60).

On the other hand, the fact that AI agents are free initiators of legal interactions is not a problem in itself, if they have been trained to comply with legal rules, but the fact that, by self-learning skills, they can change the way they function, may be. However, the self-transformation of AIs can be tolerated up to a certain point beyond which the system itself can be induced to stop working, perhaps by invalidating the digital signature system that gives official status to the legal actions it performs (Dahiyat 2021). In any case, problems linked to the identification of AI agents and the way in which they must be registered are likely to persist until widespread technological patterns are consolidated on the market. This aspect is part of the more general, albeit trivial, question of the feasibility of an 'electronic personhood', i.e. that it is not just a choice of convenience of the legislator or jurisprudence, but must also wait for certain environmental, infrastructural, and technological conditions to come about.

## **7. Limitations**

AIs make decisions and adapt to the environment with increasing levels of autonomy from their human creators and users. This is a great opportunity, but it is also problematic when things do not go as they should. These issues specifically

come into play when some of the conditions for attributing liability for the behaviour of AI systems may no longer be met by the individuals who design and/or use those systems. Often there are no short-term technical solutions to accountability gaps, but there are good reasons to bear the (social) risk of AI systems making unexpected and untraceable mistakes.

In this context, I have tried to show the weaknesses of conventional civil law liability schemes (mainly tortious one) to deal with the consequences of failures due to the proper way AIs function. The legal remedy considered was the conferral of legal personhood, justified primarily by the need to fill the resulting liability gap in the most efficient way. The various forms that legal personhood status can take were examined, as was the type of financial/asset structure on which the new legal entity should be based. The scenario of a dependent personhood status that holds AIs liable for events that are not caused by misuse, negligence, or mere design errors was considered to be more reliable. The resulting mechanism thus combines elements of negligence liability with elements akin to no-fault systems, with the provision of (limited) capital capacity to the AIs. However, a number of issues remain open and cannot be effectively addressed in this essay. Some of these issues are closely linked to the technological progress of AI systems and their actual use. Indeed, as the recent proposal for a legal framework by the European Commission suggests, the use of high-risk AIs could be severely limited or discouraged. Furthermore, technical issues remain open on how to provide an AIs with an asset. Specifically, since the same system can be sold to different users, one might ask whether it is preferable to equip the overall system with an asset or the individual services provided by the system. The two solutions require different economic assessments, with the latter one likely to be very costly.

Finally, other questions that remain unanswered have to do instead with the desirability of artificial agents with autonomous assets and independent capacity to act. In fact, it may be argued that agents capable of producing legally relevant effects in an unsupervised manner could prove, in the best-case scenario, unnecessary, in the worst-case scenario, dangerous. In these circumstances, rather than favouring legal personhood, an insurance mechanism (perhaps compulsory) might be preferred, since it has similarities in the distribution of risks and costs and does not give excessive power to the AIs. This distortion can probably be governed by constraining the active positions of legal personhood, but this hypothesis has not been sufficiently explored in this essay.

## Reference List

- AI HLEG, European Commission (2019) A definition of AI: main capabilities and disciplines. <https://ec.europa.eu/digital->

[inglemarket/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines](#)

- Andrade F, Novais P, Machado J, Neves J, (2007) Intelligent Contracting: Software Agents, Corporate Bodies And Virtual Organisations, In: Camarinha Mato L, Afsarmanesh H, Novais P, Analide C, (eds) Establishing The Foundations Of Collaborative Networks, Springer
- Andreotta AJ, (2021) The hard problem of AI rights. *AI & Society* 36: 19–32
- Atkinson R, (2019) The Case Against Taxing Robots. *Information Technology and Innovation Foundation*. <http://dx.doi.org/10.2139/ssrn.3382824>
- Ben-Ari M, Mondada F, (2018) *Elements of Robotics*, Springer, Cham
- Benson P, (2002) *Philosophy of property law*. In Coleman J, Shapiro S J, (eds.), *The Oxford Handbook of Jurisprudence & Philosophy of Law*, Oxford University Press, pp. 752—757
- Boella G, Van Der Torre L, (2007) A game-theoretic approach to normative multi-agent systems. *IEEE, Transactions on Systems, Man, and Cybernetics* 68–79
- Brożek B, Janik B, (2019) Can artificial intelligences be moral agents?. *New Ideas in Psychology* 54: 101 – 106
- Brunet P, (2019) Rights of Nature and Legal Personhood of Natural Entities in New Zeland: Making a Commons?. *Journal of Constitutional History* 38: 39-60
- Bryson JJ, Diamantis ME, Grant TD, (2017) Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence and Law* 25: 273–291
- Castelfranchi C, Dignum F, Jonker CM, Treur J (2000) *Deliberative Normative Agents: Principles and Architecture*. In: Jennings N.R., Lespérance Y. (eds) *Intelligent Agents VI. Agent Theories, Architectures, and Languages. ATAL 1999. Lecture Notes in Computer Science, vol 1757*. Springer, Berlin, Heidelberg, pp 364-378
- Chesterman S, (2020) Artificial Intelligence and The Limits Of Legal Personhood. *International & Comparative Law Quarterly* 69: 819-844
- Chopra S, White L, (2011) *A Legal Theory for Autonomous Artificial Agents*. The University of Michigan Press
- Chopra S, White L, (2009) *Artificial Agents And The Contracting Problem: A Solution Via An Agency Analysis*, University Of Illinois *Journal Of Law Technology & Policy*: 363-403
- Coeckelbergh M, (2020) Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Sci Eng Ethics* 26: 2051–2068
- Cooter R, (1991) Economic Theories of Legal Liability. *The Journal of Economic Perspectives* 5:11-30
- Dahiyat EAR, (2021) Law and software agents: Are they “Agents” by the way?. *Artificial Intelligence and Law* 29: 59–86

- Dignum F, (1999) Autonomous agents with norms, *Artificial Intelligence and Law* 7: 69–79
- Dreyfus H L, (1992) *What Computers Still Can't Do. A Critique of Artificial Reason*, MIT Press
- Freund E, (2000) *The Legal Nature of Corporations*, Batoche Books
- Gelati J, Rotolo A, Sartor G, (2002) Normative autonomy and normative co-ordination: Declarative power, representation, and mandate, Paper presented at the Workshop on the Law of Electronic Agents (LEA)
- Gibert M, Martin D, (2021) In search of the moral status of AI: why sentience is a strong argument. *AI & Society* . <https://doi.org/10.1007/s00146-021-01179-z>
- Gordon J S, (2020) *Artificial moral and legal personhood*. *AI & Society*. <https://doi.org/10.1007/s00146-020-01063-2>.
- Hansmann H, Kraakman R H, Squire RC, (2006) Law and the Rise of the Firm. *Harvard Law Review* 119: 1333-1403
- Hohfeld WN, (1913) Some Fundamental Legal Conceptions as Applied in Judicial Reasoning. *Yale Law Journal* 23: 16–59
- Huettinger M, Boyd JA, (2019) Taxation of robots – what would have been the view of Smith and Marx on it?. *International Journal of Social Economics* 47: 41–53
- Huff T, (2003) *The Rise of Early Modern Science. Islam, China and the West*, Cambridge University Press
- Pietrzykowski T (2017) The Idea of Non-personal Subjects of Law. In: Kurki V, Pietrzykowski T (eds) *Legal Personhood: Animals, Artificial Intelligence and the Unborn*, Springer, pp. 49-67
- MacCormick N, (2007) *Institutions of Law, An Essay in Legal Theory*. Oxford University Press
- Macey J R, Mitts J, (2014) Finding Order in the Morass: The Three Real Justifications for Piercing the Corporate Veil. *Cornell Law Review* 100: 99-155
- Matthias A, (2008) *Automaten als Träger von Rechten. Plädoyer für eine Gesetzänderung*. Logos Verlag, Berlin
- Mitchell T, (1997) *Machine Learning*, McGraw Hill Education
- Nevejans N, (2016) *European civil law rules in robotics [Study]*. Publications Office of the European Union. <https://doi.org/10.2861/946158>
- Oberson X, (2019) *Taxing Robots? Helping the Economy to Adapt to the Use of Artificial Intelligence*. Edward Elgar Pub
- Pietrzykowski T, (2018) *Personhood Beyond Humanism Animals, Chimeras, Autonomous Agents and the Law*. Springer
- Sartor G, (2002) Gli agenti software: nuovi soggetti di cyberdiritto in *Contratto e Impresa* 2: 57-91
- Silver D, Huang A, Maddison C, et al., (2016) Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* 529: 484–489
- Solaiman S M, (2017) Legal personhood of robots, corporations, idols and chimpanzees: a quest for legitimacy. *Artificial Intelligence and Law* 25:155–179

- Solum L B, (1992) Legal Personhood for Artificial Intelligences. North Carolina Law Review, 70: 1231-1288
- Véliz C, (2021) Moral zombies: why algorithms are not moral agents. AI & Society <https://doi.org/10.1007/s00146-021-01189-x>
- Wagner G, (2012) Comparative Tort Law. In Reimann M, Zimmermann R, (eds) The Oxford Handbook of Comparative Law (1st edn), pp. 1004 – 1040
- Zech H, (2021) Liability for AI: public policy considerations. ERA Forum 22: 147–158 <https://doi.org/10.1007/s12027-020-00648-0>
- Ziemianin K, (2021) Civil legal personhood of artificial intelligence. Future or utopia?. Internet Policy Review. <https://doi.org/10.14763/2021.2.1544>

### **Acknowledgements**

None

### **Statements and Declarations**

None