

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Apples, oranges, robots: Four misunderstandings in today's debate on the legal status of AI systems

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1728333> since 2020-02-18T18:32:04Z

Published version:

DOI:10.1098/rsta.2018.0168

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



Research

Cite this article: Pagallo U. 2018 Apples, oranges, robots: four misunderstandings in today's debate on the legal status of AI systems. *Phil. Trans. R. Soc. A* 20180168. <http://dx.doi.org/10.1098/rsta.2018.0168>

Accepted: 21 August 2018

One contribution of 9 to a theme issue 'Governing artificial intelligence: ethical, legal, and technical opportunities and challenges'.

Subject Areas:

artificial intelligence, human-computer interaction, robotics

Keywords:

accountability, artificial intelligence, legal experimentation, liability, robotics

Author for correspondence:

Ugo Pagallo

e-mail: ugo.pagallo@unito.it

Apples, oranges, robots: four misunderstandings in today's debate on the legal status of AI systems

Ugo Pagallo

Law School, University of Turin, Lungo Dora Siena 100, Torino 10153, Italy

UP, 0000-0001-7981-8849

Scholars have increasingly discussed the legal status(es) of robots and AI systems over the past three decades; however, the 2017 resolution of the EU parliament on the 'electronic personhood' of AI robots has reignited and even made current debate ideological. Against this background, the aim of the paper is twofold. First, the intent is to show how often today's discussion on the legal status(es) of AI systems leads to different kinds of misunderstanding that regard both the legal personhood of AI robots and their status as accountable agents establishing rights and obligations in contracts and business law. Second, the paper claims that whether or not the legal status of AI systems as accountable agents in civil—as opposed to criminal—law may make sense is an empirical issue, which should not be 'politicized'. Rather, a pragmatic approach seems preferable, as shown by methods of competitive federalism and legal experimentation. In the light of the classical distinction between primary rules and secondary rules of the law, examples of competitive federalism and legal experimentation aim to show how the secondary rules of the law can help us understanding what kind of primary rules we may wish for our AI robots.

This article is part of the theme issue 'Governing artificial intelligence: ethical, legal, and technical opportunities and challenges'.

1. Introduction

Scholars have increasingly discussed the legal status(es) of robots and AI systems over the past three decades

(e.g. [1,2]). However, the 2017 resolution of the EU parliament has reignited and even made current debate ideological. In the proposal, the EU institution invites the European Commission ‘to explore, analyse and consider the implications of all possible legal solutions, [including] . . . creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently’ (§59f of the document). This opinion of the EU Parliament prompted a number of censures and objections, among which the open letter of several ‘Artificial Intelligence and Robotics Experts’ in April 2018. As the letter claims, ‘the creation of a Legal Status of an “electronic person” for “autonomous”, “unpredictable” and “self-learning” robots’—as the EU Parliament suggested in its 2017 resolution—should be discarded from both a technical perspective and a normative, i.e. legal and ethical, viewpoint. The legal status for a robot, according to this latter view, can’t derive either from the ‘Natural Person model’, or from the ‘Legal Entity model’, or from the ‘Anglo-Saxon Trust model’ (§2a–c of the document).

On 25 April 2018, the EU Commission released its own document on ‘Artificial Intelligence: a European approach to boost investment and set ethical guidelines’ (doc. IP/18/3362). Remarkably, there is no reference in this document to any artificial agency, or ‘electronic personhood’, of robots and AI systems. As the press release is keen to inform us, ‘the Commission will present ethical guidelines on AI development by the end of 2018 . . . [and] by mid-2019 the Commission will also issue guidance on the interpretation of the Product Liability Directive in the light of technological developments, to ensure legal clarity for consumers and producers in case of defective products’. The overall aim is to ascertain whether and to what extent ‘artificial intelligence may raise new ethical and legal questions, related to liability or potentially biased decision-making’.

Admittedly, current trends of AI and robotics have triggered some pessimistic views. In 2015, for instance, the Future of Life Institute released an open letter addressing the challenges and threats posed by this technology: ‘Its members—and advocates, among which Bill Gates, Elon Musk, and Stephen Hawking—are concerned that as increasingly sophisticated achievements in AI accumulate—especially where they intersect with advances in autonomous robotics technology—not enough attention is being paid to safety’. A year later, the White House Office of Science and Technology Policy conducted a series of public workshops on questions of AI and policy, culminating with a report that addresses the many ethical issues related to AI, such as fairness, accountability and social justice, that should be tackled with increasing transparency. While the European Parliament’s Committee on Legal Affairs and the UK House of Commons have released similar reports on how we should prepare for the future of AI, an Industry Connections Program within the IEEE Standards Association, i.e. *The Global Initiative for Ethical Considerations in the Design of Autonomous Systems* from December 2017, presented another document, which insists on the ‘ethical concerns for autonomous and intelligent systems design’.

In light of the manifold AI robotics applications and of the multiple, and even opposite, normative views of legislators, experts and opinion makers, is there any chance to orient ourselves? Should we follow the EU Parliament’s proposal to take the electronic personhood of robots and AI systems seriously? Or, should we admit what *Nature* claimed on 25 April 2018, i.e. ‘attributing electronic personhood to robots risks misplacing moral responsibility, causal accountability and legal liability regarding their mistakes and misuses’?

The preliminary step of the analysis concerns the basic distinction between criminal law and civil law. In the former field, most of today’s debate on the impact of AI systems and robots does not regard whether such artificial agents should be granted any kind of agency, or personhood, in the legal domain. Rather, as shown by the series of meetings of the UN Certain Conventional Weapons (‘CCW’) over the past years, what is at stake regards the behaviour of AI lethal systems and robot soldiers on the battlefield, as a new source of responsibility for humans, e.g. military and political authorities. Vice versa, in the field of civil—as opposed to criminal—law, the use of AI systems and robots does not only raise new hypotheses of human responsibility for the

behaviour of some other agent in the system, such as an animal, a child or an employee. In addition to this scenario, we have to address the Gordian knot mentioned above, namely, whether such an agency of AI systems and robots could entail, and even require their legal personhood. This is the specific problem under scrutiny in this paper: once properly set the level of analysis on the legal status of AI robots, we may attempt to solve the Gordian knot without a sword. Accordingly, in order to attain such an aim, the analysis of the paper is divided into five parts.

Next, §§2 and 3 focus on the legal notions of agency and personhood: attention will be drawn to some key differences between ‘agents’ and ‘persons’, both natural and artificial. Whereas a further hypothesis on AI robots as a source of responsibility for other agents in the system, e.g. current amendments to the EU Product Liability Directive, is set aside, the intent is to show how today’s discussion on the legal status(es) of AI robots ends up with four kinds of misunderstanding. They regard the confusion between the legal personhood of AI robots and their status as accountable agents in contracts and business law.

More particularly, §4.1 examines the first sort of confusion between the apples of legal accountability and the oranges of legal personhood. Such a kind of legal reasoning can be summed up as follows: if apples, then we have oranges.

Section 4b scrutinizes the normative side of this confusion. In short, the idea is that, once AI robots are conceived of as agents in contracts and business law, thereby establishing rights and obligations in the interest of other agents, they should be treated as legal persons. In this case, we can say, if apples, then we should have oranges.

Section 4c sheds light on a variant of the previous arguments, so that, if AI robots are grasped as agents in contracts and business law, then some form of corporation follows as a result. While the aforementioned 2018 letter of AI experts to the EU Commission recalls ‘the Anglo-Saxon Trust Model’, attention should be drawn in this context to another model, that is, the Ancient Roman law idea of ‘peculium’.

Section 4d dwells on the narrative of AI robots as ‘unaccountable right violators’, or mere liability shields for humans. According to this stance, i.e. the ‘two abuses’-doctrine, granting AI robots legal personhood would be morally unnecessary and legally troublesome, for holding AI robots accountable outweighs the highly precarious moral interests that AI legal personhood might protect [3]. Although this argument has its strength, it still confuses the apples of legal accountability with the oranges of legal personhood. As a matter of fact, in the legal domain, it is not true that, if we do not have apples, neither oranges as a result.

The final part of the paper offers a map on how we may tackle the normative impact of AI robots on the civil (as opposed to the criminal) law field. Contrary to current trends of the debate on the legal status(es) of AI robots, §5 advocates a pragmatic approach. By distinguishing between the primary rules and the secondary rules of the law, §6 illustrates some implementations of this approach to the ethical and legal challenges of AI robots. Methods of legal flexibility regard the special zones, or *Tokku*, set up by the Japanese government over the past 15 years, as well as cases of legal experimentation in the field of autonomous vehicles, up to the EU regulation on data protection (‘GDPR’). By taking into account the normative challenges of AI robots, their legal status reminds us of the problem Romans used to have two millennia ago: in their view, slaves were things that nonetheless played a crucial role in trade and commerce. AI robots and enslaved persons in ancient Rome thus suggest a fruitful parallel. In both cases, individuals, e.g. human masters of both robots and enslaved persons, should not be ruined by the decisions of the latter and moreover, any counterparties of such robots and enslaved persons have to be protected when doing business, or simply interacting, with them.

The conclusions of the paper do not suggest that we should adopt any Roman solution. Still, it seems fair to admit that AI robots are increasingly affecting pillars of the law. While the EU Commission is amending the provisions of the 1985 directive on liability for defective products, it remains an open question whether the normative impact of AI robots also regards their status, either as agents or even as persons, in the legal domain. This alternative represents that which we previously dub as our Gordian knot. The overall aim of the paper is to solve this knot pragmatically, without a sword.

2. Setting the table

The next step of the analysis has to set the level of abstraction, namely, the interface that makes an investigation of some crucial aspects of the legal system possible. From a methodological viewpoint, the idea is to comprise a set of features representing the observables and variables of the analysis, the result of which provides a model for the field [4,5, pp.28–29]. From a substantial perspective, we should distinguish the analysis of the technology that is subject to legal regulation, and the set of legal notions that are at stake with matters of liability. Let us examine this twofold aspect of the problem separately in the next sections.

(a) AI robots

The definition of ‘robot’ is a tricky one: from a legal view, when some proposed to introduce a ‘robotax’ in 2017, it was far from clear what robotic applications should have been included in the list. The expansion of robotic applications for the industry and professional services has recently been complemented with the panoply of robots available out there in the field of service applications for personal and domestic use. As to the class of industrial robots, these automatically controlled, reprogrammable and multi-purpose manipulator machines are employed in a number of fields as different as refined petroleum products and nuclear fuel, textile and leather goods, communication and motor vehicles, agriculture and food products and more. As to the class of professional service robots, they are progressively employed for inspection systems, construction and demolition, logistics, professional cleaning, defence, rescue and security applications, underwater systems, mobile platforms in general use and so forth. As to the class of service apps for personal and domestic use, they include robots for home security and surveillance, for handicap assistance, or just for fun and entertainment. In the wording of the EU Agenda, ‘these robots will be bought or leased and used to provide services to individuals. They will be operated by, or interact with, untrained, or minimally trained people in everyday environments’ [6].

In light of such manifold applications, we should not miss a twofold aspect of this portrait. The first facet concerns the convergence between robotics and the Internet: avoiding the shortcomings of traditional approaches, such as on-board computers for robots, the troubles with the computing power of such machines have increasingly been addressed by connecting them to a networked repository on the Internet, allowing robots to share the information required for object recognition, navigation and task completion in the real world [7]. Although this online connectivity makes of course a lot of sense to tackle the physical world-processing tasks of robots, it presents some risks of its own, such as using these machines to perform malicious tasks under remote direction, e.g. denial-of-service attacks [8]. After all, no particular highly sophisticated self-driving car was needed to make this threat clear in August 2015, when Fiat Chrysler had to recall more than a million vehicles after a pair of hackers showed they could take over a Jeep’s digital systems over the Internet.

A second crucial aspect of current trends in robotics regards the further convergence between this domain of technological innovation and a sub-field of computer science, such as artificial intelligence (‘AI’). While spanning several disciplines, such as physics and mathematics, electronics and mechanics, neuroscience and biology, the field of robotics is progressively intertwined with advancements of AI to such an extent, that even the definition of the robot has evolved over the decades. Some argue that we are dealing with machines built basically upon the mainstream ‘sense-think-act’ paradigm of AI research [9]. Others similarly reckon that robots are machines with the ability to ‘perceive something complex and make appropriate decisions’ [10, p. 77]. Yet, also AI is quite a controversial notion. Whereas, in general terms, we can get an idea of the complexity of this domain, by looking at the list of the research fields under scrutiny in the most relevant world conferences, such as IJCAI, this paper refers to AI as the design and creation of machines that mimic (also but not only) cognitive functions that humans associate with their own intelligence, such as learning and reasoning, planning and problem solving.

The peculiarity of these smart AI and robotic systems, from a legal perspective, consists in how humans increasingly delegate crucial cognitive tasks to some of such machines, e.g. robo-traders in finance. These systems can also be unpredictable and risky, however. They increasingly gain knowledge or skills from their own interaction with the living beings inhabiting the surrounding environment, so that more complex cognitive structures emerge in the state-transition system of the artificial agent. There already are machines that can send bids, accept offers, request quotes, negotiate deals and even execute contracts. Correspondingly, this is the first time ever legal systems will hold humans responsible for what an artificial state-transition system 'decides' to do. What is—or shall be—the overall response of the law?

(b) Responsibility before the Law

The set of legal issues brought about by the behaviour of AI robots can be summed up in terms of responsibility. This perspective sheds light on the differentiation between the status of legal agents and cases in which such agents are confronted with matters of liability. On the basis of such observables and variables of the analysis on persons, agents and their responsibility, the aim is to assess three hypotheticals in this context, namely, (i) cases where the advancements of technology do not seem to affect principles and rules of today's legal systems; (ii) cases where AI and robotic technology impact on pillars of current legal frameworks and still, the use of analogy as well as the principles of the system allow lawyers to provide unequivocal solutions; and (iii) cases where there is general disagreement among experts that regard either the meaning of the terms framing the legal question, or the ways such terms are related to each other in legal reasoning, or the role of the principles that are at stake in the case.

As to the legal notion of agenthood, we should distinguish three different statuses. They regard:

- (i) Legal persons, whether natural or artificial, e.g. corporations, with rights (and duties) of their own;
- (ii) Accountable agents establishing rights and obligations in civil law, whether in the interest of other agents, or in their own interest (e.g. business by enslaved persons in Roman law); and,
- (iii) Sources of legal responsibility for other agents in the system.

As to the different types of cases where agents are confronted with legal responsibility, they regard:

- (i) Clauses of immunity (e.g. the principle of legality);
- (ii) Conditions of strict liability (e.g. no-fault responsibility of editors); and,
- (iii) Cases of responsibility for damages that depend on fault (e.g. intentional torts).

Despite crucial differences between the civil law categories for torts and damages, and the common law tradition, such conditions of immunity, faultless liability and extra-contractual responsibility cast light on how current advancement of AI technology and robotics may affect both the civil law and the common law traditions.

The next section explores which observables and variables of the analysis are relevant in today's debate on the legal status(es) of AI robots. The intent is to draw the attention to four misunderstandings that affect such debate.

3. The table

The level of abstraction set-up in the previous section suggests nine ideal-typical conditions of legal responsibility. Table 1 sums them up.

Table 1. The behaviour of AI robots before the law.

responsible AI robot	immunity	strict liability	unjust damages
as legal person	I-1	SL-1	UD-1
as accountable agent	I-2	SL-12	UD-2
as source of damage	I-3	SL-3	UD-3

Table 2. Sources of confusion in civil law.

responsible AI robot	immunity	strict liability	unjust damages
as legal person	I-1	SL-1	UD-1
as accountable agent	I-2	SL-12	UD-2

Admittedly, by taking into account the differences between legal fields, such as criminal law and civil law, conditions of legal responsibility multiply. Consider, for example, the 'Is' of table 1. Contrary to the field of civil (as opposed to criminal) law, in which analogy often plays a crucial role so as to determine individual liability, criminal law hinges on the old idea that 'everything which is not prohibited is allowed'. This principle is connected to the clause of immunity summarized, in continental Europe, with the formula of the principle of legality, i.e. 'no crime, nor punishment without a criminal law' (*nullum crimen nulla poena sine lege*). As a result, scholars have to determine whether AI robots may produce a novel generation of loopholes in the criminal law field, forcing lawmakers to intervene at both national and international levels, much as they did in the early 1990s with a new generation of computer crimes [8,11].

As stressed above in the section Introduction, however, we leave this level of analysis out in this context. While most of today's debate on the impact of robots and AI agents on the laws of war and, generally speaking, on tenets of criminal law concerns, e.g. the hypothetical 'I-3' of table 1, current discussions on how AI robots may affect pillars of civil law mainly regard cases 'I-1' and 'I-2'. Most of today's disagreement on the civil status of AI robots has in fact to do with either such AI robots as legal persons, or as accountable agents in the system, rather than simple sources of damage and, hence, of human legal responsibility. Consequently, the focus of this paper is, on the one hand, on the nine ideal-typical conditions of legal responsibility illustrated with table 1, that regard the field of civil (as opposed to criminal) law. On the other hand, we have to further restrict the focus of the paper on the first two sets of observables in table 1, thus skipping such cases as 'I-3', 'SL-3' and 'UD-3'. These latter cases regard matters of reasonable foreseeability and human negligence that affect contractual and extra-contractual interaction. Consider again the aforementioned guidance of the EU Commission on how to interpret the Product Liability Directive in light of current technological developments, e.g. machine learning techniques and algorithms capable to define or modify decision-making rules autonomously. Although national legislations may include data and information in the notion of product, it remains far from clear whether the adaptive and dynamic nature of AI robots through either machine learning techniques, or updates, or revisions, may entail or create a defect in the 'product'.

Yet, as occurs with the criminal impact of AI robots on the law, the impact of this technology as a source of responsibility for other agents in the system would require a paper of its own. Since the aim is here to illustrate the misunderstandings of today's debate on the legal status(es) of AI robots, the focus is thus restricted to the source of such misapprehension, namely, the six ideal-typical conditions of legal responsibility illustrated in table 2.

On the basis of these observables of the analysis and their variables, the time is ripe to scrutinize the misunderstandings of today's debate.

4. Apples and oranges

There are four kinds of confusion in current discussions on the legal status(es) of robots. They regard the legal status of AI robots as persons, i.e. the oranges of 'I-1', 'SL-1' and 'UD-1' of tables 1 and 2, and the legal agency of AI robots as accountable agents, namely, the apples of 'I-2', 'SL-2' and 'UD-2'. In the introduction, the arguments were summed up as follows: (i) 'if apples, then we have oranges'; (ii) 'if apples, then we should have oranges'; (iii) a variant of the first argument and, (iv) 'if we do not have apples, neither oranges as a result'.

Next sections examine each narrative separately.

(a) Confusion no. 1

The first kind of argument claims that either agency and personhood are legally equivalent or the legal agency of AI robots, e.g. the capability to act in the interest of another in the field of contracts, requires their legal personhood. Going back to the aforementioned European Parliament's proposal from February 2017, for example, it is unclear whether 'the status of electronic persons' refers to the full legal personhood of robots as proper legal 'persons', or regards their legal accountability in contracts and business law, or both. This confusion between legal persons (e.g. 'SL-1') and accountable agents (e.g. 'SL-2') reappears with certain scholars. Some claim, 'that for a computer agent to qualify as a legal agent it would need legal personhood. Both meanings of "agency" raise questions as to the desirability of legal personhood of bots' and other artificial agents such as robots [12]. More recently, this confusion between legal personhood and accountability of AI robots appears in *Of, for, and by the people* [3]. Here, the reason why legal systems should not confer legal personhood on 'purely synthetic entities' has to do with both moral reasons and the abuse of legal person status by robots and those that make them.

However, it is not necessary to resort to the example of the legal status of slaves under the ancient Roman law to show that forms of dependent or restricted legal status, such as agents in contract law acting in the interest of another, are not essentially intertwined with forms of independent legal personhood. For instance, the European Union existed for almost two decades without enjoying its own legal personhood. Rather, the EU exercised forms of strict legal agency, i.e. 'I-2', 'SL-2' and 'UD-2' of tables 1 and 2, in addition to cases in which the EU should be represented as a source of responsibility for other agents in the legal system. Therefore, scholars may discuss about different types of apple, namely, registering AI robots like corporations [13], or bestowing them with capital [14], or making the financial position of such smart machines transparent [15], without resorting to any kind of AI personhood. From the different types of apples under scrutiny in today's research, it does not follow that AI robots necessarily turn out to be an orange, thereby enjoying some form of full legal personhood.

(b) Confusion no. 2

The second kind of argument is normative, for it claims that, once AI robots are conceived of as agents in contracts and business law, then they should be treated as legal persons. As some argue, granting robots legal personhood would prevent 'the debates over slavery' that 'remind us of uncomfortable parallels with the past' and 'reflect on-going tension over humanity's role in an increasingly technologized world' [16, p. 186]. The normative ground of this reasoning rests on the reasons why legal systems grant human persons full legal personhood. As we will see below in §4.4, such reasons have to do with the moral status of humans, their intrinsic worth and capability to suffer, consciousness and so forth. As Storrs Hall holds in *Beyond AI* [17], we should accept the idea of a robot that 'will act like a moral agent in many ways', insofar as it would be 'conscious to the extent that it summarizes its actions in a unitary narrative, and ... has free will, to the extent that it weighs its future acts using a model informed by the narrative; in particular, its behaviour will be influenced by reward and punishment' [17, p. 348].

In the tradition of human rights declarations, the reference value for this kind of argument—that which I summarized years ago as ‘The Front of Robotic Liberation’ [5]—is given by the idea of dignity, e.g. Article 1 of the Universal Declaration of Human Rights (UDHR), and Protocol 13 of the European Convention on Human Rights (ECHR). However, the problem with this second kind of argument is that different types of accountability for AI robotic behaviour can be examined, regardless of whether such artificial agents are conscious, capable to suffer or experience emotions, desires, pleasures or pain. What is at stake with the legal agenthood of AI robots and their accountability in contractual and extra-contractual interaction, namely, ‘I-2’, ‘SL-2’ and ‘UD-2’ of tables 1 and 2, regards matters of efficiency in transactions and economic affairs, rather than any kind of AI robotic dignity. Advocates of the argument, according to which the legal agency of AI robots should require their legal personhood, namely, taking into account scenarios ‘I-1’, ‘SL-1’ and ‘UD-1’ of tables 1 and 2 seriously, have to preliminarily demonstrate that such AI robots possess some of the requisites that legal systems usually consider relevant, in order to grant agents their full legal personhood.

(c) Confusion no. 3

There is a variant of the previous arguments, which likens the status of AI robots to the legal personhood of corporations. In both cases, so goes the reasoning, once we admit that AI robots and corporations can establish rights and obligations in contracts and business law, then—for reasons that hinge on the efficiency of economic affairs and transactions—they should be considered as legal persons with rights and duties of their own. This idea of registering AI robots just like corporations has been popular among experts [2,18]. As claimed by other scholars, moreover, ‘existing laws might provide a potentially unexpected regulatory framework for autonomous systems’, so that we do not have to amend the current law, in order to admit that AI robots may ‘inhabit’ a company and ‘thereby gain some of the incidents of legal personality’ [19].

Yet, the idea of AI robots ‘just like corporations’ has been strongly opposed by others, including the open letter of several ‘Artificial Intelligence and Robotics Experts’ in April 2018. Granting robots the status of legal persons just like corporations would indeed be a terrible mistake. According to this narrative, if AI robots were ‘electronic persons responsible for making good any damage they may cause’, in the phrasing of the EU Parliament’s resolution from 2017, we should be ready to tackle ‘two kinds of abuse that might arise at the expense of human legal rights—humans using robots to insulate themselves from liability and robots themselves unaccountably violating human legal rights’ [3, p. 285]. In the first case, the agenthood of artificial agents could be a means to shield humans from the consequences of their conduct. Damages provoked by the behaviour and decisions of AI systems would not be upon the fat cats of Silicon Valley, because only an AI system would in fact be liable. In the second case, we should expect cases of robot insolvency: ‘money can flow out of accounts just as easily as it can flow in; once the account is depleted, the robot would effectively be unanswerable for violating human legal rights’ [3, p. 288]. Traditional punitive sanctions of the law, such as jail time for criminal insolvency, would be unavailable, unsatisfying or ineffective. On this basis, we may envisage the malfunctioning of AI robots or their manipulation that cause or fuel human wrongdoing, if not properly detected and recovered, therefore making people vulnerable to systematic recourse to such artificial agents.

Still, there are three problems with this kind of narrative on robots ‘just like corporations’. First, the corporate solution for the legal agenthood of AI robots is only one among several technical options that scholars have suggested over the past years. The list of models proposed by the ‘Artificial Intelligence and Robotics Experts’ in their letter from April 2018 is incomplete. There is no such a simple alternative between the ‘Legal Entity model’ and the ‘Anglo-Saxon Trust model’. Scholars have discussed over the past years about registries for artificial agents [2], insurance policies [20] and modern forms of the ancient Roman legal mechanism of *peculium*, namely, the sum of money or property granted by the head of the household to a slave or son-in-power [5, p. 104]. As a matter of fact, what all these examples suggest is that legal systems can properly address the challenges of the agenthood of AI robots in contracts and business law, by making

425 them accountable, without resorting to any form of corporation and hence, any kind of legal
426 personhood of AI robots. We return to this below in §§4d and 6a.

427 Second, the extent of the legal personhood of corporations varies among legal systems.
428 Contrary to the US tradition, for example, most EU companies do not enjoy their own privacy
429 rights, or their own political rights, such as freedom of speech [8]; corporations cannot be held
430 criminally responsible in the civil, as opposed to the common, law tradition [11], etc. All in all,
431 legal systems can hold the big companies of Silicon Valley accountable for what their AI systems
432 do, and yet some of these AI systems could still be conceived of as, say, ‘data processors’ pursuant
433 to Article 28 of the EU regulation on data protection. Interestingly, this is the recent idea of the
434 former Italian data protection Authority [21, pp. 124 and 173]. This scenario clearly triggers such
435 cases as the ‘I-2s’, ‘SL-2s’ and ‘UD-2s’ of tables 1 and 2.

436 Third, whether or not the legal agenthood of AI systems may make sense is an empirical issue.
437 Why, in the phrasing of *Nature*, attributing the status of, say, data processor to an AI system
438 ‘risks misplacing ... legal liability regarding their mistakes and misuses’ should not be taken as
439 a matter of principle. The issue has to do, rather, with scholarly work and legislative measures
440 that aim to prevent cases of impunity by holding AI robots accountable. While scholars have
441 suggested different kinds of strategies, as registries, insurance policies, modern forms of *peculium*
442 and the like, it is noteworthy that some institutions, as the Japanese government, have worked out
443 a way to address the legal challenges of AI robots through the creation of special zones for robotics
444 empirical testing and development, namely, a form of living lab, or *Tokku*. Significantly, the special
445 zone of Tsukuba was set up in 2011, in order to understand how AI safety governance and tax
446 regulation could be disciplined [22]. Thus, we can dismiss this part of the argument against the
447 legal agenthood of AI systems as an empirical issue on how legal systems could properly keep
448 these agents under control, and make them accountable at times.

449

450 (d) Confusion no. 4

451
452 The final argument is the trickiest: if AI robots are unaccountable right violators, then they should
453 not be granted any legal personhood as a result.

454 We may buy this argument with a pinch of salt: scholars should not lose their time debating
455 the legal personhood of AI robots, either according to the ‘Natural Personal’ model, or the
456 ‘Legal Entity’ model, lest we deal with sound accountable AI systems (which we have seen is
457 an empirical question). On this stricter basis, as conceded above in the previous section, the
458 scenario of robots as unaccountable right violators is for real. The intricacy of the interaction
459 between humans and AI systems can make it extremely difficult to ascertain what is, or should
460 be, the information content of the natural or artificial entity, as foundational to determining
461 the responsibility of individuals. Such cases of distributed responsibility that hinge on multiple
462 accumulated actions of humans and computers may lead to cases of impunity that already
463 have recommended some legal systems to adopt new forms of criminal accountability. Think
464 of the collective knowledge doctrine, the culpable corporate culture or the reactive corporate
465 fault, as ways to determine the blameworthiness of corporations and their autonomous criminal
466 liability [8].

467 However, the fallacy of the argument on the unaccountability of AI robots concerns once
468 again the confusion between apples and oranges, that is, between the legal status of AI robots
469 establishing rights and obligations, and their legal personhood. All in all, there are several
470 instances of how legal systems might grant rights of personhood, independently of matters that
471 regard accountability in contracts and torts. As to the rights of human persons, think about minors
472 and people with severe psychological illnesses, who cannot be deprived of their legal personhood
473 as espoused in certain rights despite such illnesses, or emotional and intellectual immaturity, as
474 occurs with e.g. the 1989 UN Convention on the Rights of the Child. As to the set of legal persons,
475 consider the personhood enjoyed by such non-human entities, as the Whanganui river and Te
476 Urewera national park in New Zealand, the Ganges and the Yamuna rivers in India, up to the
477 entire ecosystem in Ecuador. Therefore, in strict legal terms, the question is not about whether the

personhood of AI robots should be subject to their accountability in contracts and business law. Rather, the problem has to do with the reasons why legal systems usually grant agents their legal personhood, and whether AI robots meet such requirements. This problem brings us back to the thesis of the Front of Robotic Liberation, as introduced above in §4b.

According to the letter of the ‘Artificial Intelligence and Robotics Experts’, such an extension of the human rights model to robots ‘would be in contradiction with the Charter of Fundamental Rights of the European Union and the Convention for the Protection of Human Rights and Fundamental Freedoms’. Yet, that which advocates of the Front of Robotic Liberation claim is even more radical. Once we admit there being artificial agents capable of autonomous decisions similar in all relevant aspects to the ones humans make, the next step is to acknowledge that the legal meaning of ‘person’ and, for that matter, of constitutional rights, of dignity, of crimes, etc. will radically change. In *Bridging the Accountability Gap*, Hildebrandt *et al.* [28, pp. 558–559] warn that ‘the empirical finding that novel types of entities develop some kind of self-consciousness and become capable of intentional actions seems reasonable, as long as we keep in mind that the emergence of such entities will probably require us to rethink notions of consciousness, self-consciousness and moral agency’. This Sci-Fi scenario has two facets. The first is banal: the extension of the human rights model to robots would not only be ‘in contradiction with’, say, the 2000 EU Charter of fundamental rights. This scenario would dramatically create ‘a cleavage between human and person’ [1]. The second aspect of the problem appears more intriguing. Current AI robots lack most requisites that usually are associated with granting someone, or something, legal personhood: such artificial agents are not self-conscious; they do not possess human-like intentions, or properly suffer. Still, *pace* the ‘Artificial Intelligence and Robotics Experts’, this does not amount to say that the levels of autonomy, self-consciousness and intentionality—which arguably are insufficient to grant AI robots their full legal personhood—are inadequate to produce relevant effects in other fields of the law, e.g. the legal status of artificial agents as accountable agents in the field of contracts and business law. Otherwise, we would incur in the same kind of confusion that has been stressed apropos of, say, the ‘two abuses’-doctrine, by simply reversing the terms of such argumentation, that is, if AI robots do not meet the requisites of legal personhood, then they cannot be legal agents either.

5. A pragmatic approach

We mentioned above, in §2b, cases of general disagreement that may regard either the meaning of the terms framing the legal question, or the ways such terms are related to each other in legal reasoning, or the role of the principles that are at stake in the case. The misunderstandings of today’s debate on the legal status(es) of AI robots appear strictly connected to this latter hypothesis. Whereas advocates of the Front of Robotic Liberation support their position with the reasons why legal systems usually grant human persons full legal personhood, critics of any kind of robotic personhood or agenthood for that matter, invoke the protection of human rights. Such a polarized debate ends up with a series of Gordian knots. At times, *pace* the Front of Robotic Liberation and confusions 2 and 3 of the previous sections, the alternative is simply non-sense: should we really grant AI robots full legal personhood today?

Once we have overcome the first two confusions of today’s debate, however, a thornier set of issues follows as a result, namely, how about the legal status of AI robots as accountable agents in civil law? Should we follow the ideas of the EU Parliament or, vice versa, should we be against any agenthood of AI robots as a matter of principle, as claimed by the ‘Artificial Intelligence and Robotics Experts’ and many other scholars? Shouldn’t we at least concede the thesis of the ‘two-abuses’-doctrine, according to which the legal personhood of AI robots ultimately hinges on their role of accountable agents?

The series of Gordian knots suggests two considerations on the law as a meta-technology [5], namely, the different ways in which the law may intend to govern the process of technological innovation. Even advocates of the Front of Robotic Liberation would concede, on the one hand, that a balance should be struck between individuals claiming that they have not to be ruined by

the decisions or behaviour of their robots and AI systems and the counterparties of such machines, demanding the ability to safely interact with them. This sort of legal balance appears urgent with the spread of ‘consumer robots’, because matters of liability and extra-contractual responsibility will not only concern the giants of Silicon Valley but, increasingly, common people as well. On the other hand, the further aim which the law should have in this context concerns the advancement of technology. Legal systems should neither hinder responsible innovation nor require over-frequent revision to tackle such a progress. The wave of extremely detailed regulations and prohibitions on the use of drones by the Italian Civil Aviation Authority, i.e. ‘ENAC’, illustrates this deadlock [23]. The paradox stressed in the field of web security decades ago could indeed be extended with a pinch of salt to the Italian regulation on the use of drones: the only legal drone would be ‘one that is powered off, cast in a block of concrete and sealed in a lead-lined room with armed guards—and even then I have my doubts’ [24]. Correspondingly, we often lack enough data on the probability of events, their consequences and costs, to determine the levels of risk and thus, the amount of, say, insurance premiums and further mechanisms on which new forms of accountability for the behaviour of AI robots could hinge. As a US Navy-sponsored research admitted years ago, ‘we may paradoxically need to use the first deaths to determine the level of risk’ [25]. What is intolerable in the military sector is even more unacceptable in the civilian field. How, then, to prevent legislations that may hinder the research in AI and robotics? How to deal with their risky behaviour and even peculiar unpredictability? How should we legally regulate the future?

The short answer is ‘pragmatism’ In §4, the emphasis was on the reasons why granting AI robots legal agenthood is an empirical issue, rather than a matter of ethical thresholds on the accountability of artificial agents. Next, the aim is to illustrate how forms of legal experimentation can tackle some normative challenges of AI and robotics. The breath-taking advancements of technology in this field recommend being prepared as to how we shall rethink notions of accountability, liability and responsibility.

6. Legal experimentalism

There are two different kinds of rule, through which the law may aim to govern the process of technological innovation, e.g. the new scenarios of human–AI robot interaction (HAIRI). In accordance with Herbert Hart’s classical distinction, we should consider the primary rules of the law and its secondary rules separately [26].

(a) Primary rules of the law

The primary rules of the law aim to directly govern social and individual behaviour, both human and artificial, e.g. corporations. This kind of regulations entail four different levels of analysis, that is, (i) the regulation of human producers and designers of AI robots through law, e.g. liability norms for users of robots; (ii) the regulation of user behaviour through the design of AI robots, that is, by designing robots in such a way that unlawful actions of humans are not allowed; (iii) the regulation of the legal effects of AI robotic behaviour through the norms set up by lawmakers, e.g. the effects of AI contracts and negotiations and, (iv) the regulation of robot behaviour through design, that is, by embedding normative constraints into the design of the machine [27]. This differentiation can be complemented with further work on how the environment of HAIRI can be regulated, and the legal challenges of ‘ambient law’ [28]. Accordingly, the focus should be on the set of values, principles and norms that constitute the context in which the consequences of such regulations have to be evaluated [29].

As a case study, consider a corollary of the ‘two abuses’-doctrine—and of many views against the legal agenthood of some AI systems—such as the traditional interpretation of the role of robots and AI systems as simple tools of social interaction. This legal approach, i.e. cases ‘I-3’, ‘SL-3’ and ‘UD-3’ of table 1, mostly hinges on forms of strict liability and at times, the application of the precautionary principle. These methods of accident control aim to cut back on the scale of

the activity through the threat of physical sanctions. Yet, on the one hand, we already stressed above in §5 that such legal techniques can hinder responsible research and innovation, e.g. the drawbacks of the ENIAC regulations. On the other hand, this approach ends up in a Hegelian night where all kinds of responsibility look grey, because designers, manufacturers, operators and users of AI robots should be held accountable in accordance with the different errors of the machine and the circumstances of the case. For example, it is difficult to accept that humans should not be able to avoid the usual consequence of AI robots making a decisive mistake, e.g. the annulment of a contract, when the counterparty had to have been aware of a mistake that due to the erratic behaviour of the artificial agent, clearly concerned key elements of the agreement, such as the market price of the item or the substance of the subject-matter of a contract. Correspondingly, a three decade long debate on the legal agenthood of AI robots, namely, the ‘I-2s’, ‘SL-2s’ and ‘UD-2s’ of tables 1 and 2 illustrated above, is not a simple intellectual exercise of philosophers and dreamy experts. Rather, some of the specific solutions that have been advanced over the past years, e.g. the ‘Turing Registry’ proposed by Justice Karnow in the 1990s [2], are a response to the apparent limits of the robots-as-tools approach and a way to prevent cases of impunity.

In addition to the ‘I-2s’, ‘SL-2s’ and ‘UD-2s’ illustrated above in §4c and 4, i.e. on ‘registries’ for AI systems and robots as ‘data controllers’, consider a further example on the accountability of artificial agents in the context of public AI car sharing. In the traditional world of human drivers, many legal systems had to introduce—in addition to compulsory insurance policies—public funds for the victims of road accidents, e.g. the Italian legislative decree no. 209 from 2005. In the foreseeable world of autonomous vehicles, hypotheses of accountable AI car systems may thus make sense because a sort of digital peculium, embedded in the design of the system, can represent the smart AI counterpart to current public funds for the victims of road accidents [23]. Once clarified the level of abstraction on the variables of the legal agency, e.g. different kinds of rights and duties established by the AI system, such a legal agency should not be discarded *a priori*. Whether new forms of accountability for AI robots can properly address some drawbacks, loopholes or gaps of current legal frameworks, is indeed an empirical issue, which requires tests and experiments. Next section explores how scholars and, more importantly, lawmakers have followed this road of experimentation recently.

(b) Secondary rules of the law

In the analysis of Herbert Hart, the secondary rules of the law comprise three different types, namely, (i) rules of recognition, such as the constitution; (ii) rules of adjudication, so as to prevent e.g. conflicts of jurisdiction for extra-territorial effects of individual conduct and, (iii) rules of change. Comprehensibly, in this context, the focus is on the secondary rules of change. Leaving aside Hart’s ideas on legal theory, let us assume here the rules of change as the rules that allow to create, modify or suppress the primary rules of the system. More particularly, dealing with today’s debate on the legal status(es) of AI robots, two secondary rules of change have to be mentioned. They regard proper forms of legal experimentation and the doctrine of competitive federalism.

(c) Special zones

We can tackle some of the normative challenges of AI robots through forms of legal experimentation, as the Japanese government has worked out over the past 15 years. By creating special zones for AI and robotics empirical testing and development, the overall aim of these open labs is to set up a sort of interface, in which scientists and common people can test whether robots fulfil their task specifications in ways that are acceptable and comfortable to humans vis-à-vis the uncertainty of machine safety and legal liabilities that concern, e.g. the protection for the processing of personal data [7]. These experiments could obviously be extended, so as to strengthen our understanding of how the future of HAIRI and further smart AI robots could

turn out with some of the issues examined in the previous sections. We can collect empirical data and sufficient knowledge to make rational decisions for a number of critical issues. We can improve our understanding of how such systems may react in various contexts and satisfy human needs. We can better appreciate risks and threats brought about by possible losses of control of AI robots, so as to keep them in check. We can further develop theoretical frameworks that allow us to better appreciate the space of potential systems that avoid undesirable behaviours. Last, but not least, we can rationally address the legal aspects of this experimentation, covering many potential issues raised by the next-generation AI robots and managing such requirements, which often represent a formidable obstacle for this kind of research, as public authorizations for security reasons, formal consent for the processing and use of personal data, mechanisms of distributing risks through insurance models and authentication systems and more.

At least in the field of autonomous vehicles, some European legal systems have followed suit. Sweden has sponsored the world's first large-scale autonomous driving pilot project, in which self-driving cars use public roads in everyday driving conditions; Germany has allowed a number of tests with various levels of automation on highways, e.g. Audi's tests with an autonomously driving car on highway A9 between Ingolstadt and Nuremberg. Whereas this kind of legal experimentalism has also been implemented in other domains, such as finance, such an approach should be expanded to further fields of AI and robotics. After all, in the early 1980s, Western car producers had to learn a hard lesson when Japanese industry first began to implement robots on a large scale in their factories, acquiring strategic competitiveness by decreasing costs and increasing the quality of their products. Nowadays, it seems wise to admit that we should follow once again Japanese thinking and their policy of setting up special zones for AI robotics empirical testing and development. As stressed time and again throughout this paper, most of the issues we are dealing with in this field of technological innovation should in fact be tackled pragmatically.

But, how about the USA?

(d) Competitive federalism

Another example of secondary rules of change is given by Justice Brandeis's doctrine of experimental federalism, as espoused in *New State Ice Co. v Leibmann* (285 US 262 (1932)). The idea is to flesh out the content of the rules that shall govern social and individual behaviour through a beneficial competition among legal systems and in accordance with the principle of implementation neutrality. This is what occurs nowadays in the field of self-driving cars in the USA. On the one hand, regulations are by definition specific to that technology and yet do not favour one or more of its possible implementations. The *Federal Automated Vehicles Policy* adopted by the U.S. Department of Transportation in September 2016, illustrates this legal technique. Although regulations are specific to that technology, i.e. autonomous vehicles, there is no favouritism for one or more of its possible implementations. Even when the law sets up a particular attribute of that technology, lawmakers can draft the legal requirement in such a way that non-compliant implementations can be modified to become compliant.

On the other hand, the secondary rules of the law may help policy-makers to specify the content of the primary rules through competition between legal systems at a national level. This means that regulatory powers are first exercised by the States of the Union. More particularly, after the Nevada Governor signed a bill into law that for the first time ever authorized the use of driverless cars on public roads in June 2011, several states in the USA soon followed the example. While, as of 2016, seven states enacted laws for this kind of technology, they became 21 (plus the District of Columbia), as of February 2018. At its best possible light, this mechanism is also at work with the EU's norms on data protection, the 'GDPR'. The provisions of Article 35 on a new generation of privacy impact assessments go hand-in-hand with the powers of the supervisory authorities pursuant to Article 36 of GDPR. The idea is to pre-emptively assess the impact of new technologies on the processing of personal data, in order to minimize or prevent any kind of 'risk to the rights and freedoms of natural persons'. While the supervisory authorities of Article 36 are those of each Member State where the controller has its main establishment, a room for

690 innovation is set with the secondary rule of Article 55, because this legal mechanism of delegating
 691 powers back to states and national authorities may favour a beneficial competition among legal
 692 systems [30].

693 Admittedly, Justice Brandeis's doctrine of experimental federalism has some limits of its own.
 694 First, the non-divisibility of data and compliance costs of multi-national corporations dealing with
 695 multiple regulatory regimes can make it difficult for manufacturers catering for the international
 696 market to design in specific law abiding rules. This scenario may prompt most manufacturers to
 697 adopt and adapt themselves to the strictest standards across the board, as occurred in the case of
 698 Internet companies vis-à-vis data protection issues. Second, there are risks of fragmentation, e.g.
 699 multiple jurisdictions of national supervisory authorities in the field of EU data protection. Such
 700 risks can obviously be tackled either with technical standards, or with efforts of coordination, and
 701 yet this set of secondary rules does not guarantee *per se* a coherent interaction between multiple
 702 national legal systems and their supervisory authorities. Third, methods of legal competition
 703 between regulatory systems are provisional. They represent, after all, the way in which the
 704 secondary rules of the law should help us understanding what kind of primary rules we may
 705 wish at a federal level. Going back to the field of autonomous vehicles, this is what seems to occur
 706 in the USA, where the House of Representatives passed the Self Drive Act in September 2017. The
 707 intent is to provide a federal framework for the regulation of autonomous vehicles, regarding such
 708 aspects as 'safety assessment certifications' (§4), cyber security (§5), data privacy (§9) and more.
 709 Whether or not the US Senate will pass similar provisions in its bill is, of course, an open question.
 710 Still, in general terms, the case illustrates a wider dilemma. Methods of legal experimentalism and
 711 competitive federalism—in addition to further possible uses of the secondary rules of change—
 712 are means of legal flexibility, which should allow us to finally determine the content of the primary
 713 rules of the law. What should our conclusion be in the case of today's dilemmas on the legal
 714 status(es) of AI robots?

715 7. Conclusion

716 The paper examined a series of 'Gordian knots' that concern the polarization of today's debate
 717 on the legal status(es) of AI robots. The aim was twofold. First, the intent was to flesh out four
 718 different kinds of misunderstanding in the current debate on both the legal personhood and
 719 agenthood of AI robots and their variables. Once clarified the terms of this misunderstanding,
 720 the further aim of the paper was to stress that whether or not new forms of agenthood for AI
 721 robots can properly address drawbacks, loopholes, or gaps of the law, is an empirical issue that
 722 should be addressed pragmatically. This at the end of the day means,

- 723 (i) In the mid-term, we should skip any hypothesis of granting AI robots full legal
 724 personhood, as the EU Commission suggested in its April 2018 document on artificial
 725 intelligence;
- 726 (ii) We should take seriously into account the possibility of new forms of accountability
 727 and liability for the activities of AI robots in contracts and tort law, e.g. new forms of
 728 AI peculium for cases of complex distributed responsibility in the field of autonomous
 729 vehicles. This is what appears reasonable of the 2017 EU Parliament's resolution on the
 730 'electronic personhood of robots'; and,
- 731 (iii) We should test new forms of accountability and liability through methods of legal
 732 experimentation, widening that which has been so far the policy of several legal
 733 systems in specific domains, e.g. road traffic laws in the Japanese *Tokku* of Fukuoka
 734 since 2003.

735 On the basis of this approach, admittedly, we will not find the solution for all of the hard cases
 736 and dilemmas brought about by AI and robotics. Still, by preventing misunderstandings and
 737 the polarization of today's debate, methods of legal flexibility and pragmatic experimentation
 738 will allow us to tackle such hard cases in a rational way. In light of the probability of events,

consequences and costs of AI behaviour, we should prevent cases of legal impunity and overcome some deficiencies of current regulatory frameworks. As shown by the digital peculium for AI car sharing systems and the public funds for the victims of car accidents, new legal forms of agency for some of such AI robots may help us tackling these problems more efficiently.

Data accessibility. This article does not contain any additional data.

Competing interests. I declare I have no competing interests.

Funding. I received no funding for this study.

References

1. Solum LB. 1992 Legal personhood for artificial intelligence. *North Carolina Law Rev.* **70**, 1231–1287.
2. Karnow CEA. 1996 Liability for distributed artificial intelligence. *Berkeley Technol. Law J.* **11**, 147–183.
3. Bryson JJ, Diamantis ME, Grant TD. 2017 Of, for, and by the people: the legal lacuna of synthetic persons. *Artif. Intell. Law* **23**, 273–291. (doi:10.1007/s10506-017-9214-9)
4. Floridi L. 2008 The method of levels of abstraction. *Minds Mach.* **18**, 303–329. (doi:10.1007/s11023-008-9113-7)
5. Pagallo U. 2013 *The laws of robots: crimes, contracts, and torts*. Dordrecht, The Netherlands: Springer.
6. EU Robotics. 2013 *Robotics 2020 strategic research agenda for robotics in Europe*, draft 0v42, 11 October. Q5
7. Pagallo U. 2013 Robots in the cloud with privacy: a new threat to data protection? *Comp. Law Sec. Rev.* **29**, 501–508. (doi:10.1016/j.clsr.2013.07.012)
8. Pagallo U. 2017 AI and bad robots: the criminology of automation. In *The Routledge handbook of technology, crime and justice* (eds MR McGuire, ThJ Holt), pp. 643–653. New York, NY: Routledge.
9. Bekey GA. 2005 *Autonomous robots: from biological inspiration to implementation and control*. Cambridge, MA: The MIT Press.
10. Singer P. 2009 *Wired for war: the robotics revolution and conflict in the 21st century*. London, UK: Penguin.
11. Pagallo U, Quattrocchio S. 2018 The impact of AI on criminal law, and its twofold procedures. In *The research handbook of the law of artificial intelligence* (eds W Barfield, U Pagallo). Cheltenham, UK: Elgar.
12. Hildebrandt M. 2011 From Galatea 2.2 to Watson – and back? In *IVR world conference, August 2011*.
13. Lerouge J-F. 2000 The use of electronic agents questioned under contractual law: suggested solutions on a European and American level. *John Marshall J. Comp. Inf. Law* **18**, 403.
14. Bellia AJ. 2011 Contracting with electronic agents. *Emory Law J.* **50**, 1047–1092.
15. Sartor G. 2009 Cognitive automata and the law: electronic contracting and the intentionality of software agents. *Artif. Intell. Law* **17**: 253–290. (doi:10.1007/s10506-009-9081-0)
16. Chopra S, White LF. 2011 *A legal theory for autonomous artificial agents*. Ann Arbor: The University of Michigan Press.
17. Hall S. 2007 *Beyond AI: creating the conscience of the machine*. New York, NY: Prometheus.
18. Weitzenboeck EM. 2001 Electronic agents and the formation of contracts. *Int. J. Law Inf. Technol.* **9**, 204–234. (doi:10.1093/ijlit/9.3.204)
19. Bayern S, Burri Th, Grant ThD, Häusermann DM, Möslein F, Williams R. 2017 Company law and autonomous systems: a blueprint for lawyers, entrepreneurs, and regulators. *Hast. Sci. Technol. Law J.* **2**, 135–162.
20. Pagallo U. 2011 Killers, fridges, and slaves: a legal journey in robotics. *AI Soc.* **26**, 347–354. (doi:10.1007/s00146-010-0316-0)
21. Pizzetti F. 2018 *Intelligenza artificiale, protezione dei dati e regolazione*. Torino, Italy: Giappichelli.
22. Pagallo U. 2017 From automation to autonomous systems: a legal phenomenology with problems of accountability. In *International Joint Conferences on Artificial Intelligence Organization (IJCAI-17)*, Melbourne, Australia, pp. 17–23.

Q2
Q3
Q4

.....
rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 20180168

- 796 23. Pagallo U. 2016 Three lessons learned for intelligent transport systems that abide by
797 the law, *JusLetter IT*, 24. See [http://jusletter-it.weblaw.ch/issues/2016/24-November-2016/
798 three-lessons-learned_9251e5d324.html](http://jusletter-it.weblaw.ch/issues/2016/24-November-2016/three-lessons-learned_9251e5d324.html).
799 24. Garfinkel S, Spafford G. 1997 *Web security and commerce*. Sebastopol, CA: O'Reilly.
800 25. Lin P, Bekey G, Abney K. 2007 *Autonomous military robotics: risk, ethics, and design*. Report for
801 US Department of Navy, Office of Naval Research. San Luis Obispo, CA: Ethics + Emerging
802 Sciences Group at California Polytechnic State University.
803 26. Hart HLA. 1961 *The concept of law*. Oxford, UK: Clarendon.
804 27. Leenes R, Lucivero F. 2014 Laws on robots, laws by robots, laws in robots: regulating robot
805 behaviour by design. *Law Innov. Technol.* **6**, 193–220. (doi:10.5235/17579961.6.2.193)
806 28. Hildebrandt M, Koops B-J. 2010 The challenges of ambient law and legal protection in the
807 Q7 29. Pagallo, Durante. 2016.
808 30. Pagallo U. 2017 The legal challenges of big data: putting secondary rules first in the field of
809 EU data protection. *Eur. Data Protect. Law Rev.* **3**, 34–46. (doi:10.21552/edpl/2017/1/7)
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848