ECAI 2016
G.A. Kaminka et al. (Eds.)

doi:10.3233/978-1-61499-672-9-209

Even Angels Need the Rules: AI, Roboethics, and the Law

Ugo Pagallo[1]

Abstract.[1] Over the past years, scholars have increasingly debated over the reasons why we should, or should not, deploy specimens of AI technology, such as robots, on the battlefields, in the market, or at our homes. Amongst the moral theories that discuss what is right, or what is wrong, about a robot's behaviour, virtue ethics, rather than utilitarianism and deontologism, offers a fruitful approach to the debate. The context sensitivity and bottom-up methodology of virtue ethics fits like hand to glove with the unpredictability of robotic behaviour, for it involves a trial-and-error learning of what makes the behaviour of that robot good, or bad. However, even advocates of virtue ethics admit the limits of their approach: All in all, the more societies become complex, the less shared virtues are effective, the more we need rules on rights and duties. By reversing the Kantian idea that a nation of devils can establish a state of good citizens, if they "have understanding," we can say that even a nation of angels would need the law in order to further their coordination and collaboration. Accordingly, the aim of this paper is not only to show that a set of perfect moral agents, namely a bunch of angelic robots, need rules. Also, no single moral theory can instruct us as to how to legally bind our artificial agents through AI research and robotic programming.

1 INTRODUCTION

Over the past years the debate on "roboethics" [1, 2], and the legal aspects of robotics [3, 4], has been particularly popular among scholars. As to the technology under scrutiny, some argue that robots are machines basically built upon today's "sense-think-act" paradigm in AI research [5]. Others, as Sebastian Thrun, reckon that robots have to do with the ability of a machine to "perceive something complex and make appropriate decisions" out there [in 6, at 77]. While some others stress that robots should be able to learn and adapt to the changes of the environment, it is important to stress that robots are not a mere "out of the box" machine. As a sort of prolonged epigenetic developmental process, robots progressively 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. In addition, robots can respond to stimuli by changing the values of their properties or inner states and, furthermore, they can improve the rules through which those properties change without external stimuli. As a result, we are progressively dealing with agents, rather than simple tools of human interaction. Specimens of the same model will behave in quite different ways, according to the complexity of the context and how humans train, treat, or manage their robots. Both the behaviour and decisions of these artificial agents can thus be unpredictable and risky, hence giving rise to several normative issues.

As to the ethical and legal sides of robotics, there is an ever lasting discussion about their connection. At times, moral theories and the law simply cover different domains, or types of problem. Legal cases of faultless liability in extra-contractual obligations illustrate this point vis-à-vis the claim of virtue-ethicists that define notions of obligation, prohibition, or permission, in light of what makes life good, or bad. Most of what is morally crucial for virtue ethics is not relevant from a legal point of view: we return to this relation below in Section 3. However, contrary to current advocates of "exclusive legal positivism," we may admit that, now and then, moral theories guide the law. Consider cases of general disagreement 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 suggested by Ronald Dworkin and his followers, an option for tackling such hard cases is given by the "uniquely right answer"-thesis. According to this stance, a morally coherent narrative should grasp the law in such a way that, given the nature of the legal question and the story and background of the issue, scholars can attain the answer that best justifies or achieves the integrity of the law [7]. By identifying the principles of the system that fit with the established law, jurists could apply such principles in a way that presents the case in the best possible light.

Alternatively, some other scholars represent the hard cases of the law as a class of cases that confront us with something new and moreover, that require a reasonable compromise between many conflicting interests. Although this is of course the stance Herbert Hart made popular with his work [8], it does not follow that we have to buy any of his theoretical assumptions on, for example, the rule of recognition and the minimum content of natural law, to concede that a reasonable compromise has at times to be found in the legal domain. As previous international agreements have regulated technological advancements over the past decades in such fields as chemical, biological and nuclear weapons, or the field of computer crimes since the early 2000s, many claim that a new agreement on some of today's fields of robotics, such as robot soldiers, is necessary [9]. Regardless of the solution to the meta-disagreement on the hard cases of the law, we thus have cases in which the law needs the contribution of moral theories and a set of moral values, in order to define obligations, prohibitions, and permissions, via national statutes and international agreements, such as the Budapest Convention on computer crimes.

The stance of this paper on robots, ethics, and the law aims to explore a further kind of interaction between law and ethics. The attention is drawn here to cases in which moral theories need the contribution of the law. We may assume the ideal scenario of scholars that agree on what is right and what is wrong, what is good and what is bad, about a robot's behaviour, from an ethical point of view, and still two sets of legal issues are fated to remain

[1] Law School, University of Turin (Italy), email: ugo.pagallo@unito.it

open. By reversing the Kantian idea that a nation of devils can establish a state of good citizens, if they "have understanding" [10, at 366], we can say that even a nation of angels need some rules to further their coordination and collaboration. These rules may be interpreted either as moral norms [11], or in the sense of Hart's "secondary" legal rules, i.e. rules that allow the creation, modification, and suppression of the "primary rules" that govern people's conduct [8]. The thesis of this paper is not only that a set of perfect moral agents, namely a bunch of angelic robots, would need secondary legal rules to be good "citizens." In addition, no single moral theory can instruct us as to how we should legally bind our robots. In order to argue these theses, the paper is divided into four sections.

Next, in Section 2, attention is drawn to the debate on roboethics so as to appreciate the complexity of today's state-of-the-art. More particularly, in Section 3, the focus is on virtue ethics and how the context-sensitivity of this approach, together with its bottom-up methodology, fit like hand to glove with a pragmatic, legal approach to robotics. Section 4 illustrates two reasons why moral theories and current debate on roboethics need the support of the law. Section 4.1 scrutinizes a new generation of robotic crimes that will affect a basic tenet of the rule of law and of its continental European counterpart, the principle of legality, i.e. "no crime, nor punishment without a criminal law." Section 4.2 dwells on the creation of special, i.e. legally deregulated zones, that should allow us to test unpredictable and risky robots in open environments. The conclusions of the paper insist on how the law may help us better understand risks and threats brought on by possible losses of control of AI systems, and keep them in check. If we are fated to face some of the criminal actions sketched below in the following sections, such as e.g. the "perpetration-by-another" liability model reversed, let us address these scenarios, first, in a living lab.

2 ROBOETHICS TODAY

Scholars have increasingly discussed over the reasons why we should, or should not, deploy robots on the battlefields, in the market, or at our homes. Consider current debate on whether lethal force can be fully automated, or whether the intent to create robots that people bond with is ethically justifiable. In business law, robotic applications trading in auction markets have brought on new moral and legal dilemmas. The random-bidding strategy of these apps clarifies, or even has provoked, real life bubbles and crisis, e.g. the financial troubles of late 2009 that may have been triggered by the involvement of such artificial agents. In this context, suffice it to sum up the debate on "roboethics," or "moral machines" [12], in accordance with a twofold stance.

On the one hand, as to the strict ethical side of current discussions in the field, we should distinguish meta-ethics, applied ethics, and moral theories, such as deontologism, utilitarianism, or virtue ethics. In the field of meta-ethics, the intent is to clarify the basic concepts of the subject-matter, such as notions of right and wrong. In the field of applied ethics, scholars deal with a set of moral dilemmas arising from a specific domain, e.g. robotics. In the field of moral theories, what is at stake concerns the different ways in which we can grasp and define notions of obligation, prohibition, permission, and the like. Correspondingly, in the case of moral theories, a utilitarian would judge the action or behaviour of robots in light of their outcomes; a deontologist in connection with the intent behind such an action; a virtue-ethicist in light of what makes life good, or bad.

On the other hand, as to the technical side of the debate, there are multiple ways in which we can program our robots. This differentiation, of course, depends on the kind of moral theory we follow. However, once we agree on the content of an ethical code under a given moral theory, we can set up our robots either using deontic logic, or endorsing "principlism" and a theory of prima facie duties, or the "divine-command logic," and so forth [13]. In the case of deontic logic, the aim is to directly formalize and implement an ethical code in terms of what is obligatory, permissible, or forbidden, through an "AI-friendly"-semantics [14], and a corresponding axiomatization [15]. From the point of view of principlism, the attention is drawn to such notions as autonomy, beneficence, and the aim at doing no harm, in order to infer sets of consistent ethical rules through computational inductive logic [16]. In the case of divine-command logic, the goal is the ethical control of robotic behaviour, drawing on both the "logic of requirement" [17, 18], and modal logic [19].

In light of this panoply of approaches, both ethical and technical, we should not miss a crucial point. Regardless of today's discussions in legal theory, e.g. exclusive vs. inclusive legal positivism, it seems fair to affirm that moral theories often fall short in coping with the complexity of the legal phenomenon. Consider consequentialism, or a utilitarian stance, according to which actions, or behaviours, are judged in light of their outcomes. There are many cases in which, vice versa, "intentions" play a crucial role in the law: think of the intentional misuse of power and the reasons why a certain person committed a criminal offense, so as to evaluate the *actus reus*; the right intention of the proper authority entering into war; the intentions of the parties to a contract, or the wrongful intention that severs the link between claims of extra-contractual liability, i.e. the case of intentional torts as opposed to negligence-based responsibility and strict liability. Although we may aim to design a perfect consequentialist robot, this utilitarian approach would not prevent cases of liability for the behaviour of others in both criminal and civil law, that depend on the "intentions" of the robot.

Against this legal backdrop, some reckon that certain robots can grasp the legal terms of their behaviour and, moreover, humans could blame such machines when they do not keep their own word or when they commit some kind of offense [20, 21, 22]. Others affirm that we should be allowed to expect that a robot really means what it declares when making a contractual offer [23]. In any event, by examining, pace advocates of consequentialism, the intentions of robots, this level of abstraction deepens our understanding of, say, the good faith of humans, rather than the robots' ability to really understand what they are doing. Leaving aside the field of criminal law, to which we return below in Section 4.1, contemplate today's "contract problem" in robotics [2, 21]. Here, individuals should be held responsible for the erratic behaviour of robots, by referring the intentions of such machines to existing conventions of business and civil law, e.g. the "objective intention" of a contract. In this latter case, humans should not be able to avoid the usual consequence of robots making a decisive mistake, i.e. the annulment of a contract, when the counterparty had to have been aware of a mistake that due to the erratic behaviour of the robot, clearly concerned key elements of the agreement, such as the market price of the item or the substance of the subject-matter of that contract. Kant would agree on that.

But, reflect now on how deontologism in moral theory should address cases in which the law imposes liability regardless of the person's intentions. In addition to individuals' responsibility for the behaviour of their animals and, in most legal systems, their children, this type of faultless liability applies to most producers and users of robots. By following, e.g., Kant's theory of ethics, the aim of design should be the program of a perfect deontologist robot, so that its intentions, i.e. such cognitive states as beliefs, desires, or hopes of the artificial agent, can always be deemed as appropriate. Still, this sort of Kantian robot would not prevent the liability of its "human master," i.e. the latter's strict responsibility in cases where scholars more frequently liken robots to animals [24, 25, 26], rather than products and things. The economic rationale for this legal regime is that strict liability rules represent the best method of accident control by scaling back dangerous activities [27]. From this latter point of view, a Kantian robot, designed in accordance with the tenets of deontologism, would not be a good legal agent at all. Some times, the law does not pay any attention to intentions.

Yet, after the "consequentialist robot" and the "deontologist robot," there is a further way to conceive and design our artificial agents, i.e. according to the tenets of virtue ethics. The context sensitivity and bottom-up approach of the latter seems particularly appropriate to tackle the unpredictability and risks of robotic behaviour. As Keith Abney affirms in *Robotics, Ethical Theory, and Metaethics*, virtue ethics, rather than consequentialism, or deontologism, appears as "a more helpful approach for robots" [28]. We will explore how far this idea goes in the next section.

3 VIRTUE ROBOTS

An increasing amount of research has been devoted over the past years to the analysis of strong AI systems, trust, and security. Consider current work on the verifiability of systems that change or improve themselves, or on utility functions or decisions processes that aim to avoid that an AI system could try not to be shut down or repurposed. Likewise, reflect on further theoretical frameworks to better appreciate the space of potential systems that avoid undesirable behaviours. At the University of Stanford, an area of study has to do with "loss of control of AI systems." In the words of Eric Horvitz, "we could one day lose control of AI systems via the rise of superintelligences that do not act in accordance with human wishes [so] that such powerful systems would threaten humanity" [29]. Similar risks have been stressed by Bill Gates, Elon Musk, and Stephen Hawking. How should we address these challenges?

As mentioned above in the previous section, some reckon that virtue ethics, rather than utilitarianism, or deontologism, may help us tackling the unpredictability and risky behaviour of robots. As Abney argues, there are two reasons why this can be the case. First, this approach does not hinge on any rule-based morality but rather, draws the attention to the context sensitivity of the issues we are dealing with, namely, the disposition to act in a certain way under certain circumstances. In fact, a "proper functioning approach to evaluation appears natural: is the surgical robot operating properly in carving one's chest, or is my new robotic bandsaw dysfunctionally attempting to do the same thing?" [28]. Second, contrary to the top-down approaches of both deontologism and utilitarianism, the approach of virtue ethics is bottom-up and involves a trial-and-error learning of what makes the behaviour of a robot good, or bad. The "hybrid approach" of virtue ethics seems then particularly fruitful to tackle some of the problems with robotic behaviour, such as matters of foreseeability and due care that may trigger new cases of human negligence. The pragmatic and context sensitivity approach of virtue ethics help us indeed to determine how we should address the moral dilemmas of robotics, how we should program these machines, and test them.

However, even Abney admits the limits of this search for the virtues that properly functioning robots, given their appropriate roles, would evince. Simply put, the more societies become complex, the less shared virtues are effective, the more we need rules on rights and duties. In his words, "as the group of those dealing with robots becomes larger and more variegated, social sanctions and shared values gradually become less effective at minimizing them" [28]. Going back to the Kantian idea that even a nation of devils can establish a state of good citizens [10], we should thus admit, on the one hand, that even "virtue robots" demand rules. Yet, on the other hand, this requirement entails a twofold set of further issues. The first problem concerns the different moral rules and multiple ways in which we can embed such rules into robots. Going back to the state-of-the-art illustrated above in the previous section, should we program our robots, following a theory of prima facie duties, or the divine-command logic? Using deontic logic, or endorsing "principlism"? Should we privilege the outcomes of robotic behaviour, or judge them vis-à-vis the intent behind such actions? A mix of them?

The second problem revolves around the nature of the rules that should govern our robots. Here, we can even assume the ideal scenario of scholars that agree on the level of abstraction on what is right and what is wrong, what is good and what is bad, about a robot's behaviour. Yet, even in the case of a common ethical code under a given moral theory, a number of legal issues are fated to remain open. Whereas a set of perfect moral agents, namely a bunch of angelic robots, would still need rules to further their cooperation and collaboration [11], some of these rules are legal, rather than moral. Such rules can be grasped both in the sense of Hart's "secondary rules" that allow the creation, modification, and suppression of the primary legal rules on people's conduct [8], and as procedural rules, or of organization. The set of rules on how to produce enforceable norms at both national and international levels, along with administrative regulation at regional levels, are examples of this class of secondary rules of the law.

However, "virtue robots" also need "primary rules" that govern human and robotic behaviour in legal, rather than moral, terms. Consider cases of individual responsibility that are under a strain, such as immunity for humans bearing responsibility for the care of robots and their behaviour in the field of criminal law, or unjust damages concerning robots as a source of responsibility for other agents in the system [30]. These scenarios appear "hard," for they may spark general disagreement that does not only regard different values and principles of the normative context under examination, on which social acceptability and cohesion ultimately depend. Moreover, these cases require legal expertise to determine whether or not a loophole exists in the field, e.g. in criminal law, and hence, whether or not new primary rules should be added to the legal system.

In addition, the unpredictability of the actions or behaviour of robots, triggers an indefinite kind of cases in which we do not know where we may eventually end up. After all, the UK recorded 77 robot-provoked accidents in 2005 alone in which "people have been crushed, hit on the head, welded and even had molten

aluminium poured over them by robots" [31]. Likewise, current state of the art in technology suggests that the use of, say, unmanned aerial systems (UAS) should still be conceived as an "ultra-hazardous activity," as much as traditional aviation was considered in the 1930s [30]. Leaving aside further robotic applications, research and the breath-taking progress in AI and robotics then recommend that new levels of risk and unpredictability, e.g. cases of loss of control of AI agents, have to be taken seriously. How should we legally react before such risks and threats?

4 LAW'S EMPIRE

The Dworkinian title of this section intends to stress two different kinds of legal problem in robotics. They have to do with Hart's "primary legal rules" and their connection with the moral ones through the "secondary rules" of the law. Both problems require a particular expertise for they regard either the identification of a "loophole" in the legal system, or its inner "deadlock." At times, the behaviour of robots may of course trigger legal hard cases that bring us back to the current meta-disagreement on the hard cases of the law. We mentioned this aspect of the debate above in the introduction, e.g. the "uniquely right answer" [32] vs the "reasonable compromise"-thesis [8], so as to determine for example whether and to what extent lethal force should ever be permitted to be fully automated [9].

Here, the problem is different. It revolves around the ideal scenario of scholars that agree on what is right, or wrong, about robotic behaviour and yet, even in this case, a further set of issues is fated to remain open. These issues concern both the primary and secondary legal rules of the system that should govern the behaviour of robots, and regard either a basic tenet of the rule of law, i.e. the principle of legality, or the unpredictability of robotic behaviour. We will analyse the loopholes of the law in Section 4.1, and its deadlocks in Section 4.2. Then, the time will be ripe for the conclusions of this paper: you do not have to follow the ideas of current "exclusive legal positivism," i.e. the self-referential completeness of the law and its sources, to admit that the law has some problems of its own, also in the field of robotics.

4.1 Loopholes

The first legal problem of robotics is related to a basic tenet of the rule of law, that is summarized, in continental Europe, with the formula of the principle of legality: "no crime, nor punishment without a criminal law." Whereas certain behaviours might be deemed as morally bad, or wrong, individuals can be held criminally liable for that behaviour only on the basis of an explicit criminal norm. 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, it is likely that robots will produce a novel generation of loopholes in the criminal law field, forcing lawmakers to intervene at both national and international levels. Robot soldiers are a good example of this first kind of problem, e.g. the aforementioned question on whether lethal force should ever be permitted to be fully automated. But, consider new forms of corporate criminal liability and distributed responsibility that hinge on multiple accumulated actions of humans and computers [22, 33]. It can be extremely difficult to ascertain what is, or should be, the information content of the corporate entity as foundational to determining the responsibility of individuals. The intricacy of the interaction between humans and computers may lead to cases of impunity that have recommended some legal systems to adopt forms of criminal accountability of corporations. Think of the collective knowledge doctrine, the culpable corporate culture, or the reactive corporate fault, as ways to determine the blameworthiness of corporations and their autonomous criminal liability. Although several critical differences persist between the common law and the civil law traditions, and among the legal systems of continental Europe, we can leave aside this kind of debate, and focus on whether these forms of corporate criminal liability could be applied to the case of the artificial legal agents and the AI smart machines that are under scrutiny in this paper. Noteworthy, over the past years, several scholars have proposed new types of accountability for the behaviour of robots [23, 30, 34, 35, 36, 37], suggesting a fruitful parallelism with those legal systems that admit the autonomous criminal responsibility of corporations.

A true story helps us illustrate this new scenario: in May 2014, Vital, a robot developed by Aging Analytics UK, was appointed as a board member by the Japanese venture capital firm Deep Knowledge, in order to predict successful investments. As a press released was keen to inform us, Vital was chosen for its ability to pick up on market trends "not immediately obvious to humans," regarding decisions on therapies for age-related diseases. Drawing on the predictions of the AI machines, such trends of humans delegating crucial cognitive tasks to autonomous artificial agents will reasonably multiply in the foreseeable future. But, how about the wrong evaluation of a robot that leads to a lack of capital increase and hence, to the fraudulent bankruptcy of the corporation?

In this latter case, the alternative seems between "crimes of negligence" and the hypothesis of AI corporate liability. As to the crimes of negligence, liability depends on lack of due care, so that a reasonable person fails to guard others against foreseeable harms. The latter hinges on the traditional "natural-probable-consequence" liability model in criminal law that comprises two different types of responsibility. On the one hand, imagine either programmers, or manufacturers, or users who intend to commit a crime through their robot, but the latter deviates from the plan and commits some other offence. On the other hand, think about humans having no intent to commit a wrong but who were negligent while designing, constructing or using a robot. Although this second type of liability is trickier, most legal systems hold humans responsible even when they did not aim to commit any offense. In the view of traditional legal theory, the alleged novelty of all these cases resembles the responsibility of an owner or keeper of an animal "that is either known or presumed to be dangerous to mankind" [26].

Yet, as to the traditional crime of negligence, there is a problem: in the case of the wrong evaluation of the robot that eventually leads to the fraudulent bankruptcy of the corporation, humans could be held responsible only for the crime of bankruptcy triggered by the robot's evaluation, since the mental element requirement of fraud would be missing in the case of the human members of the board. Therefore, the criminal liability of the corporation and eventually, that of the robot would be the only way to charge someone with the crime of fraudulent bankruptcy. This scenario however means that most legal systems should amend themselves, in order to prosecute either the robot as the criminal agent of the corporation, or the corporation as such.

Further instances of new robotic offenses can be given. After all, we can apply to this context that which James Moor called the "logical malleability" of computers and so, of robots. Since the latter "can be shaped and molded to do any activity that can be characterized in terms of inputs, outputs, and connecting logical operations" [38], the only limits to the new scenarios of robotic crimes are given by human imagination. It is not so hard to envisage a world in which individuals become the innocent agent or instrument of an AI's bad decision. Certainly, by reversing the usual perspective, the scenario is not entirely new: we have full experience of hackers, viruses or trojan horses, compromising computers connected to the internet, so as to use them to perform malicious tasks under remote direction, e.g. denial-of-service attacks. Yet, what is new in the case of robots concerns their particular role of interface between the online and the offline worlds. In the internet of everything, we may envisage either powerful brain computer interfaces for robots that perceive the physiological and mental states of humans through novel Electroencephalography (EEG) filters, or robots replicating themselves, in order to specialize in infringing practices, so that no human could be held responsible for their autonomous harmful conduct. Legal systems could react either amending once again themselves, e.g. a new kind of autonomous corporate criminal liability for robots, or claiming that the principle of legality does not apply to smart machines after all. In any event, it is likely that a new general type of defence for humans, such as robotic loss of self-control, should be taken into account.

By stressing threats and risks of robotic behaviour, however, we should avert a misunderstanding. We are talking about several applications that, in the words of the UN World Robotics report from 2005, may provide "services useful to the well-being of humans" [39]. Therefore, it seems fair to affirm that the aim of the law to govern the process of technological innovation, should neither hinder it, nor require over-frequent revision to tackle such a progress. The analysis of the loopholes of today's legal systems in the field of robotics, introduces the examination of its deadlocks. Since robots are here to stay, the aim of the law should be to govern our relationships wisely.

4.2 Deadlocks

The second legal problem of robotics has to do with the unpredictability of the actions or behaviour of robots. From a legal viewpoint, the difficulty of the cases does not only regard how we should represent the web of concepts, ways of interpretation, and principles of the system that are at stake in such cases, through notions of agency, accountability, liability, burdens of proofs, responsibility, clauses of immunity, or unjust damages. Furthermore, legislators can make individuals think twice before using or producing robots, through methods of accident control that either cut back on the scale of the activity via, e.g., strict liability rules, or aim to prevent such activities through the precautionary principle [30]. The recent wave of extremely detailed regulations on the use of drones by the Italian Civil Aviation Authority, i.e. "ENAC," illustrates this deadlock [40]. How, then, to prevent legislations that may hinder the research in robotics? How to deal with their peculiar unpredictability and risky behaviour? How should we legally regulate the future?

Admittedly, the legal challenges of robotics vary in accordance with the field under examination: international law, criminal law, civil law, both in contracts and tort law, administrative law, and so forth. Some have proposed that we should register robots just like corporations in business law [34, 35, 36]; while others have recommended that we should bestow robots with capital [37], or that making the financial position of such machines transparent is a priority [23]. In the military sector, scholars and UN special rapporteurs alike have increasingly stressed over the past years, that an international agreement is needed to define the conditions of legitimacy for the employment of robot soldiers. The overall idea is that a detailed set of parameters, clauses and rules of engagement, established by an effective treaty monitoring and verification mechanisms, should allow for a determination of the locus of political and military decisions that, e.g., the increasing complexity of network-centric operations, and the miniaturization of lethal machines, can make very difficult to detect [9].

Still, in many circumstances and with most of the new generation of AI robotic applications, we have a further problem. Current default norms of legal responsibility entail a vicious circle, since the more the strict liability rules are effective, the less we can test our robots. As a result, such primary rules, e.g. the last ENAC regulation from December 2015, can indeed hinder research and development in the field. 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 insurance premiums and further mechanisms, on which new forms of accountability for the behaviour of such machines may hinge [30]. This lack of data is crucial, because the unpredictable and risky behaviour of robots affects traditional tenets of the law, such as notions of reasonable foreseeability and due care, on which people's responsibility may depend. A good example is given by how a new generation of domestic, or service, robots already impact tenets of current legal frameworks in informational privacy and data protection [3, 4, 41, 42]. Therefore, how should legal systems react?

Noteworthy, over the past 13 years, the Japanese government has worked out a way to address these issues through the creation of special zones for robotics empirical testing and development, namely, a form of living lab, or *Tokku*. After the Cabinet Office approved the world's first special zone in November 2003, covering the prefecture of Fukuoka and the city of Kitakyushu, further special zones have been established in Osaka and Gifu, Kanagawa and Tsukuba. The aim is to set up a sort of interface for robots and society, in which scientists and common people can test whether robots fulfil their task specifications in ways 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 through sensors, GPS, facial recognition apps, Wi-Fi, RFID, NFC, or QC code-based environment interaction [42]. Significantly, this approach to the risks and threats of the human-robot interaction is not only at odds with the typical formalistic and at times, pedantic interpretation of the law in Japan [43]. It is remarkable that such special zones are highly deregulated from a legal point of view. Pace the Italian ENAC, "without deregulation, the current overruled Japanese legal system will be a major obstacle to the realization of its RT [Robot Tokku] business competitiveness as well as the new safety for human-robot co-existence" [43].

So far, the legal issues addressed in the RT special zones regard road traffic laws (Fukuoka 2003), radio law (Kansai 2005), privacy protection (Kyoto 2008), safety governance and tax regulation (Tsukuba 2011), up to road traffic law in highways

(Sagami 2013). These experiments should obviously be extended, so as to further our understanding of how the future of the human-robot interaction could turn out. Some examples were illustrated above in the previous sections, such as matters of foreseeability and due care concerning human negligence, or the unpredictability of robotic behaviour that may trigger novel forms of *actus reus* in criminal law. By testing these scenarios in open, unstructured environments, the Japanese approach does not only show a pragmatic way to tackle the legal challenges of robotics. This sort of interface between strong AI robots and human societies, between present and future, also allows us to better comprehend risks and threats brought on by possible losses of control of AI systems, so as to keep the latter in check.

5 CONCLUSIONS

There are three different ways in which we can grasp the theses and title of this paper, "even angels need the rules." The first way directly concerns today's debate in roboethics. From a moral point of view, we can say that even a set of perfect agents, namely a bunch of angelic robots, need rules to further their cooperation and collaboration. Even advocates of virtue ethics concede this point [28]. As illustrated above in Sections 2 and 3, these rules can be interpreted either as moral ones [11], or as secondary legal rules [8].

A second way to interpret the title of the paper involves Hart's primary rules and more particularly, that which Section 4.1 presented as the loopholes of the law. The legal impact of robotics affects all the sectors in the legal field and still, is especially relevant in criminal law, where analogy should not help tackling the impact of robotics. Although moral theories play a role in this context, so as to either find out the uniquely right answer (Dworkin), or a reasonable compromise (Hart), moral theories do not instruct us as to whether and to what extent we are confronted with a legal loophole and hence, whether or not new legal rules should be added to the system. This is a question that appears crucial for today's debate on roboethics and still, goes beyond the expertise of robo-ethicists. Is there any loophole in the legal system?

The third way to inflect the title regards the unpredictability and risky behaviour of robots that have been stressed time and again in this paper. Whilst no single moral theory can tackle the complexity of the law and instruct us as to how we should legally bind our robots, the law itself is confronted with that which Section 4.2 summed up as the practical and theoretical "deadlocks" of today's legal systems. Lawmakers often make individuals think twice before using or producing robots, through methods of accident control that either cut back on the scale of the activity, or aim to stop these activities at all. Here, what the law adds to the current debate in roboethics has to do with the definition of specific secondary legal rules that should allow us to understand what kind of primary legal rules we may need. The creation of legally de-regulated, or special, zones for robotics appears a smart way to overcome such deadlocks and to further theoretical frameworks with which we should better appreciate the space of potential systems that avoid undesirable behaviours. By testing the human-robot interaction outside laboratories, i.e. in open or unstructured areas, we can improve our understanding of how these artificial agents may react in various contexts and satisfy human needs. Also, we can rationally manage the legal aspects of this experimentation, covering many potential issues raised by the next-generation robots and tackling those requirements that often represent a formidable obstacle for this kind of research, such as public authorisations for security reasons, formal consent for the processing and use of personal data, mechanisms of distributing risk through insurance models and authentication systems, and the like. This is the set of secondary legal rules with which to strengthen our comprehension of the type of primary legal rules we need in order to govern our robots. At the end of the day, this sort of legal de-regulation also offers a fruitful way to deepen our understanding of some moral dilemmas in the field.

REFERENCES

[1] G. Veruggio, Euron Roboethics Roadmap, *Proceedings Euron Roboethics Atelier*, February 27th-March 3rd, Genoa, Italy (2006)

[2] J. J. Bryson, The Meaning of the EPSRC Principles of Robotics, *The AISB Workshop on Principles of Robotics,* 4 April 2016.

[3] U. Pagallo, Killers, Fridges, and Slaves: A Legal Journey in Robotics, *AI & Society*, **26(4)**, 347-354, 2011.

[4] RoboLaw, Guidelines on Regulating Robotics. EU Project on Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics, 22 September 2014.

[5] G. A. Bekey, *Autonomous Robots: From Biological Inspiration to Implementation and Control*, The MIT Press, Cambridge, Mass. & London, 2005.

[6] P. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, Penguin, London, 2009.

[7] R. Dworkin, *A Matter of Principle*, Oxford University Press, Oxford, 1985.

[8] H. L. A. Hart, *The Concept of Law*, Oxford University Press, Oxford, 1961.

[9] U. Pagallo, 'Robots of Just War: A Legal Perspective', *Philosophy and Technology*, **24(3)**, 307-323, (2011).

[10] I. Kant, Perpetual Peace, *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*, trans. by M. Gregor, vol. 8, Cambridge University Press, Cambridge, 1999.

[11] L. Floridi, 'Distributed Morality in an Information Society', *Science and Engineering Ethics*, **19(3)**, 727-743, (2013).

[12] W. Wallach and C. Allen, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, New York, 2009.

[13] S. Bringsjord and J. Taylor, The Divine-Command Approach to Robot Ethics, in *Robot Ethics: The Ethical and Social Implications of Robotics*, pp. 85-108, edited by P. Lin, K. Abney and G. A. Bekey, The MIT Press, Cambridge, Mass., 2014.

[14] J. Horty, *Agency and Deontic Logic*, Oxford University Press, New York, 2001.

[15] Y. Murakami, Utilitarian Deontic Logic, *Proceedings of the Fifth International Conference on Advances in Modal Logic*, pp. 288-302, edited by R. Schmidt et al. AiML, Manchester UK, 2004.

[16] M. Anderson and S. Leigh Anderson, Ethical Healthcare Agents, *Advanced Computational Intelligence Paradigms in Healthcare*, pp. 233-257, edited by M. Sordo et al. Springer, Berlin, 2008.

[17] R. Chisholm, Practical Reason and the Logic of Requirement, *Practical Reason*, 1-17, edited by S. Koerner, Basil Blackwell, Oxford, 1974.

[18] Ph. Quinn, *Divine Commands and Moral Requirements*, Oxford University Press, Oxford, 1978.

[19] C. I. Lewis and C. H. Langford, *Symbolic Logic*, Dover, New York, 1959.

[20] S. J. Hall, *Beyond AI: Creating the Conscience of the Machine*, Prometheus, New York, 2007.

[21] S. Chopra and L. F. White, *A Legal Theory for Autonomous Artificial Agents*, The University of Michigan Press, Ann Arbor, 2011.

[22] G. Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, Dordrecht, 2015.
[23] G. Sartor, 'Cognitive Automata and the Law: Electronic Contracting and the Intentionality of Software Agents', *Artificial Intelligence and Law*, **17(4)**, 253-290, (2009).
[24] B. Latour, *Reassembling the Social: an Introduction to Actor-Network-Theory*, Oxford University Press, Oxford, 2005.
[25] D. McFarland, *Guilty Robots, Happy Dogs: The Question of Alien Minds*, Oxford University Press, New York, 2008.
[26] J. Davis, 'The (common) Laws of Man over (civilian) Vehicles Unmanned', *Journal of Law, Information and Science*, **21(2)**, 10.5778/JLIS.2011.21.Davis.1., (2011).
[27] R. Posner, *Economic Analysis of Law*. Little Brown, Boston, 1973.
[28] K. Abney, Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed, *Robot Ethics: The Ethical and Social Implications of Robotics*, 35-52, edited by P. Lin, K. Abney and G. A. Bekey, The MIT Press, Cambridge, Mass., 2014.
[29] E. Horvitz, One-Hundred Year Study of Artificial Intelligence: Reactions and Framing. White Paper. Stanford University, at https://stanford.app.box.com/s/266hrhww2l3gjoy9euar,(2014).
[30] U. Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts*, Springer, Dordrecht, 2013.
[31] R. Noack, A Robot Killed a Factory Worker in Germany. So Who Should Go on Trial?, *The Washington Post*, 2 July 2015.
[32] R. Dworkin, *Law's Empire*, Harvard University Press, Cambridge, Mass., 1986.
[33] P. M. Freitas, F. Andrade and P. Novais, Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible, *AI Approaches to the Complexity of Legal Systems*, 145-156, edited by Pompeu Casanovas et al., Springer, Dordrecht, 2014.
[34] C. E. A. Karnow, 'Liability for Distributed Artificial Intelligence', *Berkeley Technology and Law Journal*, **11**, 147-183, (1996).
[35] J.-F. Lerouge, 'The Use of Electronic Agents Questioned under Contractual Law: Suggested Solutions on a European and American Level', *The John Marshall Journal of Computer and Information Law*, **18**, 403, (2000).
[36] E. M. Weitzenboeck, 'Electronic Agents and the Formation of Contracts', *International Journal of Law and Information Technology*, **9(3)**, 204-234, (2001) .
[37] A. J. Bellia, 'Contracting with Electronic Agents', *Emory Law Journal*, **50**, 1047-1092, (2001).
[38] J. Moor, 'What Is Computer Ethics?', *Metaphilosophy*, **16(4)**, 266-275, (1985).
[39] UN World Robotics, Statistics, Market Analysis, Forecasts, Case Studies and Profitability of Robot Investment, edited by the UN Economic Commission for Europe and co-authored by the International Federation of Robotics, UN Publication, Geneva (Switzerland), 2005.
[40] ENAC, Remoted Piloted Aerial Vehicles Regulation, issue No. 2 dated 16 July 2015. Revision 1 dated 21 December 2015.
[41] U. Pagallo, 'Robots in the Cloud with Privacy: A New Threat to Data Protection?', *Computer Law & Security Review*, **29(5)**, 501-508, (2013).
[42] U. Pagallo, The Impact of Domestic Robots on Privacy and Data Protection, and the Troubles with Legal Regulation by Design, *Data Protection on the Move*, 387-410, edited by S. Gutwirth, R. Leenes, and P. de Hert, Springer, Dordrecht, 2016.
[43] Y.-S. Weng, Y. Sugahara, K. Hashimoto and A. Takanishi, 'Intersection of "Tokku" Special Zone, Robots, and the Law: A Case Study on Legal Impacts to Humanoid Robots', *International Journal of Social Robotics*, published online on February 13, (2015).