From Automation to Autonomous Systems: A Legal Phenomenology with Problems of Accountability

Ugo Pagallo
University of Turin, Italy, Law School
ugo.pagallo@unito.it

Abstract

Over the past decades a considerable amount of work has been devoted to the notion of autonomy and the intelligence of robots and of AI systems: depending on the application, several standards on the "levels of automation" have been proposed. Although current AI systems may have the intelligence of a fridge, or of a toaster, some of such autonomous systems have already challenged basic pillars of society and the law, e.g. whether lethal force should ever be permitted to be "fully automated." The aim of this paper is to show that the normative challenges of AI entail different types of accountability that go hand-in-hand with choices of technological dependence, delegation of cognitive tasks, and trust. The stronger the social cohesion is, the higher the risks that can be socially accepted through the normative assessment of the not fully predictable consequences of tasks and decisions entrusted to AI systems and artificial agents.

1 Introduction

The paper offers a concise phenomenology on how automation and the development of artificial intelligence ("AI")-systems have affected pillars of the law. The purpose is threefold: in Section 3, focus is on the notion of autonomy as a source of misunderstanding in today's multidisciplinary debate. Section 4 sheds light on the normative challenges of technology and whether or not automation and AI may trigger loopholes in the legal field. Section 5 draws the attention to different types of legal accountability that go hand-in-hand with choices of technological dependence, delegation of cognitive tasks, and trust. The relation between law and technology should be grasped as the interaction between competing regulatory systems that not only may reinforce or contend against each other, but against further regulatory systems, such as the forces of the market and of social norms. Against this framework, the conclusion of the analysis aims to elucidate the proper attitude with which we should address such a competition, together with the compromises that, at times, are necessary in the legal domain.

2 A Legal Phenomenology

Current debate on cognitive automata in the form of software agents and AI systems can be traced back to the seminal remarks of German scholars on automation and the law in the late 1800s. This co-evolution of technology and legal systems has known so far three major steps: they concern the ancestors of today's debate, the dawn of AI, and the turning point of the latter in the early 2000s. Each of these steps is examined separately in the following sections.

2.1 Ancestors

Scholars started examining the legal impact of automation, e.g. automatic vending machines, since the last decade of the 1800s. Think of Günther's *Das Automatenrecht* (1892), Schels' *Der strafrechtliche Schutz des Automen* (1897), Schiller's *Rechtsverhältnisse des Automen* and Ertel's *Der Automatenmissbrauch und seine Charakterisierung als Delikt*, both from 1898, up to Neumond's *Der Automat* in 1899. From a civil—as opposed to the criminal—law viewpoint, what initially was at stake concerned the role that the will of the parties to a contract could play with automation. From a criminal law perspective, the discussion revolved around whether and to what extent the process of automation could have produced a novel generation of loopholes in the criminal law field, forcing lawmakers to intervene. Contrary to the field of civil law, in which analogy often plays a crucial role so as to determine individual liability, individuals can be held criminally liable for their behaviour only on the basis of an explicit criminal norm. The principle is related to a basic tenet of the rule of law, which is summarized, in continental Europe, with the formula of the principle of legality: "no crime, nor punishment without a criminal law."

More than a century later, this kind of debate is still going on. Reflect on the EU data protection regulation n. 679 from 2016, or "GDPR," according to which individuals have a right to explanation that derives from the notification duties of the data controllers, in order to provide the data subjects with all the information necessary to ensure fair and transparent processing. On the one hand, pursuant to Articles 13(2)(f) and 14(2)(g) of the GDPR, this information regards "the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject." On the other hand, according to Article 22(1), "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." Leaving aside the tricky meaning of some formulas of this article—that refer either to general clauses, such as that which "significantly" may affect the data subject, or to vague notions on what "produces legal effects"—it remains unclear how to interpret the levels of human involvement in the decision-making process, e.g. how to interpret the formula of "a decision based solely on automated processing." Would the result of an automated processing that is not actively assessed by any human, but is formally attributed to them, fall beyond the scope of Article 22(1)?

There is however a crucial difference between the legal debate on automation from the 1890s and current discussions on automated processing. The technological leap concerns the "logic involved" in such automated processing. The latter increasingly regards a particular class of algorithms that either augment or replace analysis and decision-making by humans, as occurs with the discipline of machine learning, i.e. algorithms capable to define or modify decision-making rules autonomously. The second step of our phenomenology has thus to do with the field of AI and more particularly, with the crucial shift from automation to artificial autonomy.

2.2 The Dawn of AI

There are two ways in which we can appreciate the impact of AI on current legal systems. The first way concerns the grandfather of current work on AI and the law—namely, the great German polymath Gottfried W. Leibniz—and his dream to make legal reasoning and enforcement automatic through the use of combinatorial analysis, probability calculus, and binary arithmetic [Pagallo, 2005 and 2016a]. From this stance, it follows that the law can be conceived of as a rich test bed and important application field for logic-based AI research that regards: (i) applications of logic to the representation of legal regulations, where the legal conclusions follow from that representation as a matter of deduction; (ii) applications of logic to legal rules that are not just applied but are the object of reasoning and discourse; and, (iii) both interpretative reasoning in light of the facts of a case, and evidential reasoning to establish such facts (as in, say, applying statutory rules in unforeseen circumstances). On the basis of “formal models of legal procedure and of multi-agent interaction in legal proceedings,” the overall idea is that the scope of logic is widening from deduction to information flow, argumentation and interaction [Prakken and Sartor, 2015].

On the other hand, attention should be drawn to the original plan of AI, namely, the design and setting 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. Here, focus is on how AI systems and apps, such as smart software agents, AI robots, or other autonomous systems may affect the requirements and functions of the law, i.e. what the law is supposed to be (requirements), and what it is called to do (functions). Admittedly, after the birth of AI in the 1950s and the grand expectations of both the founding fathers and leaders of this kind of research, we had to wait for more than Simon's "25 years," or Minski's "current generation," in order to shift from simple automation to robust autonomous systems [Simon, 1965; Minski, 1967]. Although we still have not got either machines that are capable of doing any work men can do, or the solution for the problem of creating proper artificial intelligence, we are increasingly dealing with systems that 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 AI application. Rather than simple machines that can operate or control a process mechanically, in other words, we are progressively interacting with proper artificial agents.

2.3 The Turning Point

Over the past decade, the feasibility, importance, and scalability of current AI technologies have gone hand-in-hand with the rapid progress of "four self-reinforcing trends," that concern the improvement of more sophisticated statistical and probabilistic methods, the increasing availability of large amount of data and of cheap, enormous computational power, up to the transformation of places and spaces into IT-friendly environments, e.g. smart cities and domotics [Cath et al., 2016]. What this huge transformation means from a legal point of view, can be summed up here with two crucial points.

The first remark brings us back to Leibniz and more specifically, to G.W.F. Hegel's critique of his work. As stressed in the *Lectures on the History of Philosophy* (part III, C, 1), Hegel admired Leibniz's principles of "individuality" and of "indistinguishability," and yet he claimed, "this is an artificial system, which is founded on a category of understanding, that of the absoluteness of abstract individuality" [Hegel, ed. 1892-6]. In simpler terms, what Leibniz missed was the impact of his own ideas and projects on the real world, that is, how making legal reasoning automatic would have changed both requirements and functions of the law. Going back to current advancements in the field of AI, the same remark holds true. The more AI attains its own aim to create systems that function in smart ways, the less our world remains unaffected, the more we have to pay attention to the trends and effects of this profound transformation. We return to this stance below in Section 4.

The second crucial point regards a popular topic of today's debate as well as a major source of misunderstanding among scholars and policy makers, namely, the aforementioned shift from automation to the autonomy of AI systems. There are indeed three different ways in which we can grasp the notion of autonomy. The first meaning has to be distinguished from further notions of adaptability and interactivity [Allen et al., 2000]. From this perspective, an AI system can be conceived of as an autonomous system when it modifies its inner states or properties without external stimuli, thereby exerting control over its actions without any direct intervention of humans [Floridi and Sanders, 2004]. Such a property can of course complement both the interactivity and adaptability of the system. In the first case, the system perceives its environment and responds to stimuli by changing the values of its own properties or inner states. In the second case, an AI system is adaptable, when it can improve the rules through which its own properties or inner states change. Taken together, such criteria single out why we should refer to an AI system as an "agent," rather than a simple tool of human interaction.

The second meaning of autonomy summarizes the previous properties of what constitutes the notion of agency with a single word. Consider the UK Ministry of Defense's Joint Doctrine Note on "unmanned aircraft systems" from March 2011. The notion of autonomy there is connected to a system "capable of understanding higher level intent and direction." Along these lines, there can be different levels of autonomy, and of automation. For example, in the *Federal Automated Vehicles Policy* adopted by the U.S. Department of Transportation in September 2016, the latter distinguishes six levels based on who does what, and when, i.e. from level 0 in which human drivers do everything to level 5 in which "the automated system can perform all driving tasks, under all conditions that a human driver could perform them."

The final meaning of autonomy is the more controversial, since it likens AI autonomy to human autonomy, including at times "free will" and "moral agency" as used to describe human decision-making. Although this stance appears quite science fictional, we find it time and again in current debate, either for philosophical reasons, or for political motives, e.g. the "Campaign to Stop Killer Robots" launched in April 2013 by both a prominent non-governmental organization, such as *Human Rights Watch*, and the Harvard International Human Rights Clinic. Leaving aside in this context the political understanding of this latter debate [Burri, 2016], let us examine the anthropomorphist meaning of AI autonomy separately in the next section.

3 Are You Autonomous, Aren't You?

A considerable amount of work has been devoted over the past years to studying the normative challenges of fully autonomous AI systems and robots, that is, according to the third meaning of autonomy, as illustrated above in the previous section. Scholars have examined such scenarios as AI systems developing an interest in self-preservation, or harming humans so as to benefit themselves, or accessing quantum computers capable of cracking the most sophisticated control code systems. In addition, scholars have discussed whether and to what extent AI systems should enjoy their own rights, such as the right to be free from pain and suffering, the right to free speech, and so forth. This is what, some years ago, I dubbed as the theses of the Front of Robotic Liberation [Pagallo, 2013].

Among the most committed advocates of a new generation of AI crimes and rights, suffice it to mention the work of Gabriel Hallevy. In his view, AI technology "has the capability of fulfilling the awareness requirements in criminal law" [Hallevy, 2015], together with "the mental element requirements of both intent offenses and recklessness offenses." This not only means that AI systems can be either liable as direct perpetrators of criminal offenses, or responsible for crimes of negligence, or on strict liability basis, and so on. Moreover, the general defense of loss of self-control, insanity, intoxication, or factual and legal mistakes, could protect such artificial agents. Once the mental element requirement is fulfilled in the case of a robot, there would be no reason why the general purposes of punishment and sentencing, i.e. retribution and deterrence, rehabilitation and incapacitation, down to capital penalty, should not be applied to AI machines.

Although we may buy Lawrence Solum's argument that "one cannot, on conceptual grounds, rule out in advance the possibility that AIs should be given the rights of constitutional personhood" [Solum, 1992]—or, for that matter, that the traditional paraphernalia of criminal lawyers could be properly extended to the regulation of AI systems—there are two problems with this kind of stance. On the one hand, if we admit there being AI machines capable of autonomous decisions similar in all relevant aspects to the ones humans make, the next step would be to acknowledge that the legal meaning of "person," along with that of crimes of intent, of negligence, of strict liability, etc., will radically change. Even Solum admits that, "given this change in form of life, our concept of a person may change in a way that creates a cleavage between human and person." Likewise, in [Hildebrandt et al., 2010], they 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." At the end of the day, nobody knows to where this scenario may lead. For instance, would a strong AI robotic lawyer accept Hallevy's argument that "evil is not part of the components of criminal liability" [Hallevy, 2015]? What if the AI agent, rather than an advocate of current exclusive legal positivism, is a follower of the natural law tradition?

On the other hand, at the risk of being lambasted for reactionary anthropocentrism, we should admit another kind of priority. Rather than debating Sci-Fi scenarios of artificial agents endowed with human-like properties, such as free will or moral sense, attention should be drawn to how the behaviour of some AI systems already falls within the loopholes of the law, provoking a new generation of hard cases that necessitate the intervention of lawmakers at both national and international levels. We do not have to wait, in other words, for any top level of automation, e.g. level 5 of the US *Federal Automated Vehicles Policy*, in order to admit that some current AI systems have already challenged basic pillars of society and the law. Lawyers must be pragmatic, after all.

4 Accountable AI

As mentioned above in Section 2.1, we should distinguish between criminal law and civil law, in order to appreciate the normative challenges of AI technology and their impact on current legal systems. Whereas, in the field of civil law, analogy can play a crucial role in order to define matters of accountability and personal responsibility, criminal liability can be established only on the basis of an explicit norm. Next sections illustrate this very difference through some examples in the field of criminal law (section 4.1), and of civil law (section 4.2). Then, Section 5 examines how legislators and policy makers should address the normative challenges of AI.

4.1 Criminal Accountability

As the grandfather of legal automation and AI & law, Wilhelm Leibniz, used to say, "every mind has a horizon in respect to its present intellectual capacity but not in respect to its future intellectual capacity" (quoted by [Coudert 1995]). In light of Leibniz's wisdom, I risk here a projection. The scenario is inspired by a true story: 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. This type of liability 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 AI system, 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 an AI application. 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" [Davis, 2011].

Yet, as to the traditional crime of negligence, there is a problem: in the case of the wrong evaluation of the AI system that eventually leads to the fraudulent bankruptcy of the corporation, humans could be held responsible only for the crime of bankruptcy triggered by the AI system'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 AI agent 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.

4.2 Civil Accountability

The current traditional interpretation of AI systems' behaviors in the field of civil law conceives such systems as simple tools of social interaction. Whereas, in criminal law, the accountability for bad AI behavior is typically imposed on individuals who voluntarily commit a wrong prohibited by law, in the field of civil law we should further distinguish between contracts and torts. In contracts, civil accountability mostly regards compensation to those affected by the harmful behavior of a counterparty through the AI system; in tort law, payment follows from obligations between private persons usually imposed by the state to compensate for damage provoked by AI wrongdoing. In both cases, there are some problems.

In tort law, consider the European Directive 85/374/EEC on liability for defective products. Although national legislation implementing the directive may include data and information in the notion of product, it remains far from clear whether and to what extent the adaptive and dynamic nature of AI through machine learning techniques, updates, or revisions, may entail or create a defect in the "product." In addition, we should take into account the scenario of AI systems that functioned adequately but produced a harmful outcome on the basis of erroneous data, or bad inputs. Again, it is far from clear who should be held accountable, i.e. either the producer or manufacturer of the AI system, or the supplier of the data, such as the internet operator that failed providing connectivity, or both. However, data suppliers cannot be considered so far as "producers" in the sense of the European directive; and moreover, the ecosystem behind an AI behavior can be so complex that may severely affect the ability of lawyers to sever the chain of liability through notions of legal causation and fault [Karnow, 1996]. No surprise then, that the EU Commission started the process for the amendment of the aforementioned directive in September 2016.

As to the field of contracts, the AI-as-tools approach means that rights and obligations established by the artificial agent (AA) directly bind the human principal (P), since all the acts of AA are considered as acts of P. In addition, P cannot evade liability by claiming either she did not intend to conclude such a contract or AA made a decisive mistake. In this latter case, e.g. in case of the erratic behaviour of AA, what P can do is to claim damages against the designer and producer of AA. According to the mechanism of the burden of proof, P will have to demonstrate that AA was defective and that such defect existed while AA was under the manufacturer's control; and furthermore, the defect was the proximate cause of the injuries suffered by P. Still, it is difficult to accept that rights and obligations established by AAs would be directly conferred upon humans, because the principal wanted the specific content, or agreement, of the contract made by AA. Rather, rights and obligations are conferred onto humans because they delegate to AA the authority to act on their behalf. Whereas the traditional approach ends up in a Hegelian night where all kinds of responsibility look grey, it thus seems necessary to amend today's rules of the law, so that operators and users of robots should be held accountable in accordance with the different errors of the machine and the circumstances of the case. For example, 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 behavior 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. In general terms, the aim should be to strike a balance between individuals claiming that they should not be ruined by the decisions or behaviour of their AAs and the counterparties of such machines, demanding the ability to safely interact with them. This is the balance that has been aimed at by several scholars [Allen and Widdison, 1996; Weitzenboeck, 2001; Barfield, 2005; Sartor, 2009; Pagallo, 2013; etc.).

However, this latter approach faces a major problem, that is, the lack of data that depends either on current default rules of legal responsibility, or on the impossibility to test the AI systems in unstructured environments. Since such problems have to be examined vis-à-vis the different ways in which legislators aim to govern the field of technological research and development, these issues are deepened separately in the next section.

5 The Challenges of Regulation

The previous sections have illustrated some cases brought on by AI technologies that, sooner or later, will induce national and international legislators to intervene in the fields of criminal and civil law. All in all, legislators are confronted with three different kinds of challenge. They concern (i) the specific features of AI technology; (ii) the competition between regulatory systems; and, (iii) how to address such challenges at a meta-regulatory level.

As to the AI features, from a legal viewpoint, the most salient aspect of this technology has to do with the unpredictable behavior of AI systems that may hinge on machine learning techniques, or on the complexity of the ecosystem behind them, etc. Legislators have so far tackled this scenario 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. Current default norms of legal responsibility can entail however a vicious circle, since the more the strict liability rules are effective, the less we can test our AI systems, the more such rules may hinder research and development in the field [Pagallo, 2016b]. The recent wave of extremely detailed regulations on the use of drones by the Italian Civil Aviation Authority, i.e. "ENAC," illustrates this deadlock. As a result, 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.

Yet, to make things even more intricate, we have to pay attention to the competition between regulatory systems, such as the forces of the market, or of social norms. Every regulatory system claims to govern social behaviour by its own means, and can even render the claim of another regulatory system inadequate, or superfluous. Reflect on all the cases in which the legal intent to regulate the process of technological innovation has failed, e.g. the EU e-money directive 46 from 2000. Soon after its implementation, further forms of online payments, such as PayPal, forced the Bruxelles legislators to intervene, amending themselves with the new directive 110 from 2009. Regardless of the scenario we consider, however, such a competition between regulatory systems does not take place in a normative vacuum, but is structured by the presence of values and principles [Pagallo and Durante, 2016]. In addition to the technological and legal standards of the field, focus should thus be on the epistemic standards, i.e. ways to understand the informational reality of AI systems, and both the social standards that enable users to trust such AI systems and evaluate the quality of the skills regardless of whether or not these skills meet social needs. Accordingly, policy makers and legislators should keep in mind the degree of social acceptability and cohesion that affect their own decisions, that is, whether or not a shared set of values and principles exist. This bifurcation is critical, because it tells us something new about the normative challenges of AI from a meta-regulatory point of view. There is a fundamental choice that has to be taken, regarding the delegation of decisions to autonomous AI systems. The political resolution does not only depend on the degree of predictability and reliability of the AI decisions. Rather, the issue may revolve around the degree of social agreement, or disagreement, that characterize the normative context under examination. How, then, should lawmakers address the interaction between competitive regulatory systems? How can we prevent legislations that may hinder the research in AI? How should we deal with the peculiar unpredictability and risky behavior of some AI systems? How should we legally regulate the future?

Luckily enough, there are multiple legal techniques with which we can properly address this set of challenges. Suffice it to mention here three of them. First, focus should be on 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 individual behavior through a beneficial competition among legal systems. This is what occurs nowadays in the field of self-driving cars in the US, where seven states have enacted their own laws for this kind of technology. At its best possible light, the same policy will be at work with the EU regulation in the field of data protection [Pagallo, 2017b].

Second, attention should be drawn to the principle of implementation neutrality, according to which regulations are by definition specific to that technology and yet do not favor one or more of its possible implementations. The *Federal Automated Vehicles Policy* of the U.S. Department of Transportation—which was mentioned above in Section 2.3—illustrates this legal technique. Although regulations are by definition specific to that technology, e.g. autonomous vehicles, there is no favoritism 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.

Third, legislators can adopt forms of legal experimentation. For example, over the past 14 years, the Japanese government has worked out a way to address the normative challenges of robotics through the creation of special zones for their empirical testing and development, namely, a form of living lab, or *Tokku* [Weng et al, 2015]. Likewise, in the field of autonomous vehicles, several EU countries have endorsed this kind of approach: 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. In general terms, these forms of experimentation through lawfully de-regulated special zones represent the legal basis on which to collect empirical data and sufficient knowledge to make rational decisions for a number of critical issues. We can improve our understanding of how AI systems may react in various contexts and satisfy human needs. We can better appreciate risks and threats brought on by possible losses of control of AI systems, 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 behaviors. In addition, we can rationally address the legal aspects of this experimentation, covering many potential issues raised by the next-generation AI systems 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.

Of course, some of these legal techniques can interact and reinforce each other [Pagallo, 2017c]. More importantly, they represent a mechanism of legal flexibility that allows us to address the interaction between regulatory systems wisely. At the end of the day, 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 manage such progress [Pagallo, 2016c].

6 Conclusions

The paper has offered a concise phenomenology on automation, autonomous AI systems, and the law, in order to determine to what extent this field of technological innovation has already affected today's legal systems. By discarding Sci-Fi scenarios in Section 3, Section 4 dwelt on three ways in which the behavior of some AI systems falls within the loopholes of the law. Correspondingly, legislators have to tackle three different kinds of challenge that were examined in Section 5. They concern the specific features of AI technology, the interaction between competitive regulatory systems, and how lawmakers can address such challenges at a meta-regulatory level. Some legal techniques, such as the rules of experimental federalism or the implementation neutrality-approach, can offer mechanisms of legal flexibility that allow us to properly address such challenges.

All in all, there is a fundamental choice that has to be taken, regarding the growing delegation of decisions to autonomous AI systems and a myriad of smart artificial agents. The political resolution does not only depend on the degree of predictability and reliability of the AI decisions. The issue may revolve around the degree of social agreement, or disagreement, that characterize the normative context under examination. Tasks entrusted to AI systems, autonomous robots, or software agents, affect assets and interests that does not only affect the degree of "social acceptability" that concerns the risk inherent in the delegation process. The technical and legal decisions on how tasks delegated to AI systems may impact on assets and human interests, can also regard values and principles that ground those assets and interests.

Matters of technological dependence and the corresponding grade of delegation and autonomy have thus to be comprehended in accordance with the degree of social cohesion that exists in the normative context in which the consequences of tasks and decisions delegated to AI systems are evaluated. The definition of legal standards, as a means that allows agents to communicate and interact, has to take into account the interplay with further regulatory systems and the extent to which social cohesion is affected by technology, e.g. unemployment triggered by robotics [Floridi, 2017]. The stronger the social cohesion is, the higher a risk in the delegation process that can be socially accepted through the normative assessment of not fully predictable consequences of entrusted tasks and decisions to AI systems and autonomous artificial agents. Since AI systems are here to stay, the aim of the law should be to wisely govern our mutual relationships. The mechanisms of legal flexibility illustrated in this paper show how this can be possible.

References

[Allen and Widdison, 1996] Tom Allen and Robin Widdison. Can computers make contracts? *Harvard Journal of Law & Technology* 9(1): 26–52, 1996.

[Allen et al., 2000] Colin Allen, Gary Varner and Jason Zinser. Prolegomena to any future artificial moral agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 12: 251–261, 2000.

[Barfield, 2005] Woodrow Barfield. Issues of law for software agents within virtual environments, *Presence* 14(6): 741–748, 2005.

[Burri, 2016] Thomas Burri. The Politics of Robotic Autonomy. *European Journal of Risk Regulation*, 7(2): 341-360, 2016.

[Cath et al., 2016] Corinne Cath, Sandra Wachter, Brent Mittelstadt, Mariarosaria Taddeo and Luciano Floridi. *Artificial Intelligence and the 'Good Society': the US, EU, and UK Approach*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2906249, December 2016 (last accessed 19 April 2017).

[Coudert, 1995] Allison P. Coudert. *Leibniz and the Kabbalah*. Kluwer Academic, Boston-London 1995.

[Davis, 2011] Jim Davis. The (common) Laws of Man over (civilian) Vehicles Unmanned, *Journal of Law, Information and Science*, 21(2): 166-179, 2011.

[Floridi, 2017] Luciano Floridi. Robots, Jobs, Taxes, and Responsibilities, *Philosophy & Technology*, 30: 1-4, 2017.

[Floridi and Sanders, 2004] Luciano Floridi and Jeff Sanders. On the Morality of Artificial Agents, *Minds and Machines*, 14(3): 349–379, 2004.

[Hallevy, 2015] Gabriel Hallevy. *Liability for Crimes Involving Artificial Intelligence Systems*. Springer, Dordrecht, 2015.

[Hegel, 1892-6] *Lectures on the History of Philosophy*. English translation by E.S. Haldane, available at https://www.marxists.org/reference/archive/hegel/works/hp/hpconten.htm (last accessed 19 April 2017).

[Hildebrandt et al, 2010] Mireille Hildebrandt, Bert-Jaap Koops and David-Olivier Jaquet-Chiffelle. Bridging the accountability gap: Rights for new entities in the information society? *Minnesota Journal of Law, Science & Technology*, 11(2): 497–561, 2010.

[Karnow, 1996] Curtis E. A. Karnow. Liability for Distributed Artificial Intelligence, *Berkeley Technology and Law Journal*, 11: 147-183, 1996.

[Minski, 1967] Marvin Minski. *Computation: Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, N.J., 1965.

[Pagallo, 2005] Ugo Pagallo. *Introduzione alla filosofia digitale: da Leibniz a Chaitin*. Giappichelli, Torino, 2005.

[Pagallo, 2013] Ugo Pagallo. *The Laws of Robots: Crimes, Contracts, and Torts*. Springer, Dordrecht, 2013.

[Pagallo, 2016a] Ugo Pagallo. *Leibniz: una breve biografia intellettuale*. Kluwer, Padova, 2016.

[Pagallo, 2016b] Ugo Pagallo. Even Angels Need the Rules: AI, Roboethics, and the Law. In Gal A Kaminka et al (eds), *ECAI 2016. Frontiers in Artificial Intelligence and Applications*, pp. 209-215. IOS Press, Amsterdam 2016.

[Pagallo, 2016c] Ugo Pagallo. Three Lessons Learned for Intelligent Transport Systems that Abide by the Law, *JusLetter IT*, 24, November 2016. Available at http://jusletter-it.weblaw.ch/issues/2016/24-November-2016/three-lessons-learne_9251e5d324.html.

[Pagallo, 2017a] Ugo Pagallo. The Legal Challenges of Big Data: Putting Secondary Rules First in the Field of EU Data Protection, *European Data Protection Law Review*, 3(1): 36-46, 2017.

[Pagallo, 2017b] Ugo Pagallo. When Morals Ain't Enough: Robots, Ethics, and the Rules of the Law, *Minds and Machines*, January 2017.

[Pagallo and Durante, 2016] Ugo Pagallo and Massimo Durante. The Pros and Cons of Legal Automation and its Governance, *European Journal of Risk Regulation*, 7(2): 323-334, 2016.

[Prakken and Sartor, 2015] Henry Prakken and Giovanni Sartor. Law and Logic: A Review from an Argumentation Perspective. *Artificial Intelligence* 227: 214-245, 2015.

[Sartor, 2009] Giovanni Sartor. Cognitive automata and the law: Electronic contracting and the intentionality of software agents, *Artificial Intelligence and Law* 17(4): 253–290, 2009.

[Simon, 1965] Herbert Simon. *The Shape of Automation for Men and Management*. Harper & Row, New York, 1965.

[Solum, 1992] Lawrence B. Solum. Legal personhood for artificial intelligence. *North Carolina Law Review*, 70: 1231–1287, 1992.

[Weng et al., 2015] Yueh-Hsuan Weng, Yusuke Sugahara, Kenji Hashimoto and Atsuo Takanishi. Intersection of "Tokku" Special Zone, Robots, and the Law: A Case Study on Legal Impacts to Humanoid Robots, *International Journal of Social Robotics*, 7(5): 841-857, 2015.

[Weitzenboeck, 2001] Emily Mary Weitzenboeck. Electronic Agents and the Formation of Contracts, *International Journal of Law and Information Technology*, 9(3): 204-234, 2001.