Working paper, comments welcome
09 March 2017

LAW AND REGULATION OF ARTIFICIAL INTELLIGENCE AND ROBOTS: CONCEPTUAL FRAMEWORK AND NORMATIVE IMPLICATIONS

Nicolas Petit*

INTRODUCTION

Discussions over the law and regulation of Artificial Intelligence ("AI") and robots are all the rage as early applications are introduced in society. In computer science, concerns that "*overly rigid regulations might stifle innovation*",[1] have fueled proposals to create regimes of selective immunity for research on intelligent machines.[2] At the same time, ethical arguments have prompted calls for an all-out ban on research in relation to lethal automated weapons ("LAWs").[3] And some writers claim that robots will become so important to mankind that "*a new branch of the law*" is needed, "*to grant their race and its individual members the benefits of legal protection*", much like the international community did, or tried to, with the environment.[4]

Within the legal profession, the discussion focuses less on public policy issues. In each specialist field of the law, experts discuss whether AIs and robots call for changes to the rules that govern their discipline.[5] Take copyright law. A typical discussion is whether the rise of robots-created works necessitates to modify the statutes that grant protection to "*the author*", which is manifestly inapplicable for robots.[6] More generally, this approach can be seen in the

* Professor of Law, University of Liege (ULg) ; Research Professor, University of South Australia (UniSA) ; Senior Scholar, International Center for Law and Economics ("ICLE"). Comments can be sent at Nicolas.petit@ulg.ac.be

[1] *The Economist*, You, Robot? 01 September 2012.

[2] R. Calo, "Open robotics." *Maryland Law Review* 70.3 (2011): 101-142.

[3] Autonomous Weapons: An Open Letter from AI & Robotics Researchers, 28 July 2015, http://futureoflife.org/open-letter-autonomous-weapons/: "*Starting a military AI arms race is a bad idea, and should be prevented by a ban on offensive autonomous weapons beyond meaningful human control*".

[4] D. Levy, *Robots Unlimited, Life in a Virtual Age* (2006) by A K Peters, Ltd., p.397.

[5] This piecemeal, disciplinary approach, has already been followed in the past in relation to the rise of the Internet, 3D-printing or big data.

[6] D. Levy, *supra* at 396-397.

mushrooming of books, publications and symposiums on "*artificial intelligence and "[liability; criminal; tax; privacy; etc.] law*".

Despite its pragmatism, the flaws of the disciplinary approach of the regulation of AIs and robots are obvious. The rules devised by experimented lawyers in one silo of the legal system may not consider spill-over effects on other disciplines.[7] Take for instance the inconsistency between an innovation-adverse rule of strict liability on AI programmers on the one hand and an innovation-friendly legal framework that encourages computer scientists to work on AI through the allocation of subsidies, intellectual property ("IP") rights and tax benefits on the other hand. Or consider a statute that confers dignity rights to robots. As part of such rights, a prohibition of torture would prevent to test how a robot reacts in stressful circumstances, and in turn undermine experimental research on the design of safer robots.

Similarly, many pitfall affects attempts to build a specific law of AI and robots. In a famous paper titled "*Cyberspace and the Law of the Horse*", Judge Easterbrook criticized attempts to build an *ex novo* law of cyberspace, noting that error in legislation is common, especially when technology is emerging at rapid pace.[8]

Fortunately, the early scholarship has treaded with caution. Comprehensive approaches to the law and regulation of AIs and robots have been proposed, possibly in the hope to assist decision makers. Two dominant routes have been followed. The first is *legalistic*. It consists in starting from the legal system, and proceed by drawing lists of legal fields or issues affected by AIs and robots: liability, privacy, cyber security, etc.[9] The second is *technological*. The point here is to envision legal issues from the bottom-up standpoint of each class of technological application: driverless cars, social robots, exoskeletons, etc.[10]

---

[7] As Oliver Holmes once noted, "*the life of the law has not been logic: it has been experience*". *See* O.W. Holmes, Jr., Lecture I in *The Common Law* (1881) .

[8] F.H. Easterbrook, "Cyberspace and the Law of the Horse." *U. Chi. Legal F.* (1996): 207-216.

[9] C. Leroux, et al. "Suggestion for a green paper on legal issues in robotics." Contribution to Deliverable D 3.1 (2012) (reviewing six disciplines of the law likely to be affected); M. De Cock Buning, L. Belder and R. de Bruin, "Mapping the Legal Framework for the introduction into Society of Robots as Autonomous Intelligent Systems" available at: http://www.caaai.eu/wp-content/uploads/2012/08/Mapping-L_N-fw-for-AIS.pdf (who build six relevant case scenarios which correspond to disciplines where legal issues will arise (standardization, IP, legal capacity, liability, privacy and AIs generated works)).

[10] E. Palmerini, et al. "RoboLaw: Towards a European framework for robotics regulation." *Robotics and Autonomous Systems* 86 (2016): 78-85 ; E. Palmerini, "The interplay between law and technology, or the

Much of the emerging literature on the law and regulation of AIs and robots has followed either/or both routes.

A third often-mentioned approach distinguishes between robot ethics and robolaw. Robot ethics follow the "*garbage in/garbage out*" precept that governs computer sciences. It insists on "*Norms [that] can be directly incorporated into technology in the sense that a command and the compliance to it are imbued in the technology itself*",[11] like in Isaac Asimov's three laws of robotics[12].[13] Robolaw focuses on external norms that govern the operation of AI and robots once introduced in society[14].[15]

Like all problem identification exercises, the three abovementioned approaches are necessarily uncertain. Moreover, many tabled regulatory initiatives seem *either* to occur without an explicit methodological foundation *or* to mix a bit of all approaches. The 2017 European Parliament resolution on Civil law rules on robotics is a case in point.[16] This rich catalogue of ambitious regulatory proposals does not articulate the principles used to select issues and to propose remedies. As a result of this methodological indeterminacy, the resolution displays a number of inconsistencies. For instance, it seems to suggest that it is

---

RoboLaw project in context." *Law and Technology. The Challenge of Regulating Technological Developments* (2013): 7-24.

[11] E. Palmerini, et al., "RoboLaw. Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics", (2014) available at: http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf

[12] I. Asimov, "Runaround", *Astounding Science Fiction* 1942.

[13] A more comprehensive definition of robot ethics includes the regulation of technology development, including by regulating "*human ethics to be followed by the researchers*". *See* D. Levy, *supra* at 400. Another possible distinction is between robot ethics and machine ethics, which considers the engineering of advanced robots that are not present yet in our society. *See* http://io9.gizmodo.com/5936838/who-should-pay-when-your-robot-breaks-the-law.

[14] The robot ethics/robolaw distinction separates between acting on the robot design itself or acting at the execution stage. But it also seems to display sociological ramifications. Where Robolaw seems primarily a field populated by lawyers, robot ethics also boast philosophers, ethicist, psychologists, computer scientists and novelists.

[15] Yet, this approach is less about eliciting law and regulation options, than about implementing law and regulation prescriptions.

[16] *See* European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

inappropriate to impute liability on humans for acts of autonomous robots, but at the same time calls for compulsory insurance on users.

This is why this paper seeks to articulate an alternative methodology for the law and regulation of AIs and robots in a context of technological emergence.[17] The perspective taken is that of a benevolent social planner. The issue in discussion consists in localizing the regulation needs generated by the introduction of AIs and robots in society. By regulation, I mean "*State intervention into the economy by making and applying legal rules*".[18]

To be even more concrete, this paper attempts to address the following normative question: should a benevolent social planner adopt specific rules and institutions for AIs and robots or instead, should the resolution of issues can be left to Hume's three "*fundamental laws of nature*",[19] namely ordinary rules on property and liability,[20] contract laws and the courts system.[21] I call this the basic legal structure. To explore that question, this paper applies a public interest framework. I choose this approach not because of an ideological belief that the public interest is the best guide of the regulatory hand. Instead, my choice is governed by logic. Public interest regulation is accepted by both consequentialists and liberal paternalists as one of the least intrusive regulatory methodologies. It is therefore the first next step if one starts from the assumption of a world governed by Hume's laws of nature.

---

[17] In May 2016, the Committee on Legal Affairs of the European Parliament has issued a Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). This report invites other bodies to evaluate the implications of radical proposals, such as creating a specific legal status for robots or compulsory insurance schemes for producers or owners of robots.

[18] B. Morgan, and K. Yeung, *An introduction to law and regulation: text and materials*, Cambridge University Press, 2007.

[19] D. Hume, *A treatise of human nature*, Courier Corporation, 2012 ("*stability of possession, of its transference by consent, and of the performance of promises*").

[20] I do not discuss whether existing regulatory schemes that go beyond the basic legal structure are appropriate for AIs, for this is a largely empirical question. However, as Bertolini observes, law already exists – for instance existing product liability schemes – that can be readily applied to robotics. A. Bertolini, "Robots as products: the case for a realistic analysis of robotic applications and liability rules." *Law, Innovation and Technology* 5.2 (2013): 214-247.

[21] Whilst our approach certainly cuts through the existence of many existing rules that can potentially apply to AI beyond the basic legal structure (for instance, product liability), our focus on the basic legal structure as the fallback system avoids the pitfall of assuming that a "*disembodied free market*" could solve all issues. On this, see R.A. Epstein, "Can Technological Innovation Survive Government Regulation." *Harv. JL & Pub. Pol'y* 36 (2013): 87 ("*At bottom, the proper inquiry never poses the stark choice of regulation versus no regulation*".

This paper proceeds as follows. In section 1, it reviews several regulation approaches proposed in the literature (I). In section 2, it discusses the regulatory trade-offs, namely the threats and opportunities created by the introduction of regulation in a context of technological evolution (II). In Section 3, it looks at the specific area of liability to discuss some law and regulation options (III). In section 4, it proposes a possible methodology for the law and regulation of AIs and robots (IV).

Before moving to the analysis of those issues, two last remarks are in order. *First*, I discuss both AIs and robots under the same intellectual roof, even though I am aware of the differences between those two technological fields. I do this not only because I conjecture a degree of convergence between both technologies, but primarily because intelligent machines in soft or hard envelopes have the ability to "*act upon the world*".[22]

Second, some may consider that this paper's subject of inquiry falls beyond the scope of legal scholarship. Yet, the idea that a lawyer shall not discuss normative issues rests on a mistaken perspective on the legal system (that I have fought time and time again). The lawyer occupies a decentralized place in society. His work is by essence empirical. At bottom, the lawyer argues, assesses and comments "*cases*". This position makes him a firsthand witness of social, economic, moral and technological *facts*, and a useful adviser to the social planner. His reasoning, often qualitative, is by no means less scientific – and in fact very complementary – to that of other professions who routinely advise the political technocracy, like positive economics which occasionally work on formal models based on assumptions remote from real facts.[23] And the *trial* of facts and law is a process which is useful – though not ideal – to reveal imperfect information, which is precisely what central governance institutions lack.

[22] R. Calo, "Robotics and the Lessons of Cyberlaw." *Cal. L. Rev.* 103 (2015): 513-564.

[23] R. Coase, "The conduct of economics: the example of Fisher Body and General Motors." *Journal of Economics and Management Strategy* 15.2 (2006): 255-278.

II. OVERVIEW OF PROPOSED REGULATION APPROACHES

A. LEGALISTIC APPROACH

Under a legalistic approach, the social planner attempts to visualize how existing laws apply to an AI or a robotic application.[24] In other words, the question is therefore whether AI and robotic application is caught by existing legal rules. And the gist of the discussion is about frictional rule-implementation cases. The analysis is either conducted vertically within a specific field of the law by specialist lawyers, or horizontally across several fields of the law by generalist lawyers. In both cases, the assessment is conducted on a disciplinary basis. The disciplines commonly looked at in the scholarship are: product safety (including cyber security) and liability, consumer protection, intellectual property, labour law, privacy, civil liability, criminal liability, legal personhood, insurance and tax law.

To focus on a specific example, let us look at IP. A recurring discussion is whether AI-created inventions can be protected under IP laws (and if this is the case, who is their owner)[25].[26] Consider an AI-created song. In the area of copyright law, there is a requirement of "*originality*". Across the globe, courts insist that the work exhibits a "*modicum of creativity*", reflects the "*author's intellectual creation*" or constitutes the exercise of "*non-mechanical, non-trivial skill and judgment*".[27] A debate today exists on whether the originality requirement prevents the allocation of copyrights to intelligent machines. And a proxy is that a selfie taken by a monkey has been deemed unsusceptible of copyright protection.[28]

---

[24] This can be understood with a little green men metaphor. An alien from a distant planet sets foot on Earth. Most of Earth's clothing factories produce ready-to-wear suits for humans in calibrated sizes. Are human suits fitting, must they be stretched, adjusted, refitted? Or shall humans leave the alien naked?

[25] By AI-created invention, I mean an invention fully brought to existence by the AI, without human assistance.

[26] C. Leroux et.al., *supra* at 37: "*Above mentioned legal IP systems are based on the fact that computers are inert tools, so that current intellectual property regimes usually only apply to humans or legal persons creations and not to creations coming from computers or inert tools. However, artificial technologies have advanced rapidly to the point that intelligent agents do not assist humans in the creation of works, but generate them autonomously. Thus, intelligent agents are capable of creativity*".

[27] E.F. Judge, and D. Gervais. "Of silos and constellations: Comparing notions of originality in copyright law." *Cardozo Arts & Ent. LJ* 27 (2009): 375-408.

[28] Though primarily on grounds of lack of authorship. *See*, on this case, J. Jowitt, "Monkey See, Monkey Sue: Gewirth's Principle of Generic Consistency and Rights for Non-Human Agents." *Trinity CL Rev*. 19 (2016): 71-96.

Similarly, in the area of patent law, an innovation is protected on condition that it involves an "*inventive*" step. As a rule, inventiveness means non-obvious to a person skilled in the art. In layman's term, a non-obvious discovery is one that is unexpected. But where to set the benchmarks for "*non-obviousness*" and "*skill in the art*", when one contemplates the introduction of AIs capable of "*recursive self-improvement*".[29] What is not obvious to a man skilled in the art may be trivially evident for a super intelligent machine.

Whilst it is not my intention here to dwell on those specific issues, I want to stress that the legalistic approach has advantages and disadvantages. Its upside is to trigger a process of prospective and retrospective discussion of the social choices embedded in legal arrangements (hardly ever a bad thing). Its downside is informational. Lawyers remote from technological fields conjecture frictional rule-implementation cases with imperfect comprehension of the underlying science. With this, mistakes are inevitable. Lawyer may elaborate wrong hypotheses, and in turn create new legal problems. Suppose that lawyers posit (perhaps unconsciously) that the path to cure Alzheimer's disease is computational, and that the emergence of AIs is likely to accelerate society's progress towards a treatment. On that basis, the social planner adopts strong patent protection regimes for AIs and computational inventions in this field. But consider now that the technology frontier for a computational cure of Alzheimer's disease is in reality unapproachable. And suppose that scientists believe more in biotechnological solutions like digital brain emulation (hereafter emulated minds or "EMs"). Given the strong patent protection granted to computational AI technologies, sunk investments flow in the (wrong) direction, and not towards the more promising EMs. This, in turn, fuels scepticism against the patent system, and spurs initiatives to weaken (abolish) patent protection altogether. At the end of the line, this chills R&D investments into a wide array of beneficial technologies, both AIs and EMs.

To put the point differently, AIs and robots are emergences. They generate entirely unpredictable sets of issues. Reasoning from existing rules is likely to generate blind spots. This is because our rules are abstract commands designed on the basis of specific representations of the state of the world, and its trajectories.[30] Too much reliance on the

[29] For use of this term, see N. Bostrom, *Superintelligence, Paths, Dangers, Strategies*, Oxford University Press, 2014.

[30] Take again the metaphor of the Alien. His physiological condition may be different from ours. He may not need a suit at all. Instead, his physiological needs could be entirely different: for instance, he may be over sensitive to noise, and need soundproof helmets to cover its ears.

legalistic approach may undermine the necessary development of novel legal fields in a context of emergences.[31] Let me take two examples to illustrate the point. *First*, consider that an AI creates a revolutionary time-travel application. Users can now travel through centuries in virtual (or actual) reality, and interact with ancient parents. Should not we introduce a *law of time travel*, given the risk of adverse psychological effects when users land in the real world?

Second, assume that an AI ends-up dominating the world.[32] If this dire prediction ever came true, must society introduce a "*law of humans*" which affords minority rights to humans and protects our species from intelligent machines? We certainly cannot see this necessity today because all our laws are based on the non-secular postulate that human conscience is special, and makes us superior to machines, animals and other entities. As a result of this cultural prior, our laws are essentially more about how humans treat non-humans (and in particular machines), than about how non-humans (and in particular machines) treat humans.

B. TECHNOLOGICAL APPROACH

Under the technological approach, the social planner attempts to understand if AIs and robots generate issues which deserve treatment by law and regulation. The method followed consists in identifying classes of AI and robotic applications and to gauge legal needs from there. The technological approach does not treat law as a given, and it is therefore more inclined to envision a lawless AI world. Moreover, it is more technology friendly than the legalistic approach.[33]

The Stanford *Artificial Intelligence and Life in 2030* report (the "Stanford Report") provides an illustration.[34] The report purports to highlight how AI applications bring "*specific changes*

---

[31] J.M. Balkin, The Path of Robotics Law. *California Law Review* 6 (2015): 45-60. ("*The new technology disrupts the existing scene of regulation, leading various actors to scramble over how the technology will and should be used. As people scramble and contend with each other over the technology, they innovate—not only technologically, but also socially, economically, and legally—leading to new problems for law. Instead of saying that law is responding to essential features of new technology, it might be better to say that social struggles over the use of new technology are being inserted into existing features of law, disrupting expectations about how to categorize situations*").

[32] Bostrom refers to this as the singleton scenario. *See* N. Bostrom, *supra* at Chapter V.

[33] To use again our little green men metaphor, the technological approach is comparable to a tailor-made suit factory. The exercise consists in designing cloth that suits the alien.

[34] Artificial Intelligence and Life in 2030, One Hundred Year Study on Artificial Intelligence, Report of the 2015 Study Panel, September 2016.

affecting the everyday lives of the millions of people who inhabit them". The focus is placed on eight applications domains where AI has or is likely to have the greatest impact: "*transportation, service robots, healthcare, education, low-resource communities, public safety and security, employment and workplace, home/service robots, and entertainment*". From this, the Stanford Report enumerates nine broad categories of legal and policy […] that AIs tend to raise: privacy, innovation policy, civil liability, criminal liability, agency, certification, labor, taxation and politics.

The Stanford Report displays commonalities, but also discrepancies with the legalistic approach.[35] Some topics that were absent, irrelevant or subjacent in the legalistic approach are prominent in the technological approach. This is the case of the legal arrangements governing certification (*eg*, professional licensing requirements), taxation (*eg*, how automated compliance reduces revenue from infringements to the law) and politics (*eg*, voting and deliberation processes). And in the technological approach, the legal issues are framed as general topics that cut through several legal fields. Innovation policy is a heading under which liability law, freedom of speech and patent law are discussed.

Moreover, the technological approach is more opened to questions of robot ethics and *ex ante* legal design.[36] A commonly discussed problem is the trolley dilemma,[37] which envisions how to code an AI to sort decisional options that entail human killing.[38] This is distinct from the legalistic approach which focuses predominantly on legal arrangements for *ex post* frictional cases. One reason for this difference may be that technology-untrained lawyers are less comfortable discussing how to turn legal rules into computer code.

The technological approach seems again more technology-optimistic than the legalistic approach. With this, the discussion often revolves around whether and how law and regulation risks stifling innovation incentives, research and development investments and inventive activity.

---

[35] For instance, the Stanford Report stresses that AI is very relevant in relation to "*regulation*" without though discarding its relevance for other sources of law like common law, federal law, local statutes or ordinances.

[36] As explained above, this refers to the *ex ante* coding of legal rules in AIs and robots at the design stage.

[37] www.robolaw.euE. Palmerini (2014), *supra*. ("*train conductors are in a situation in which they have to decide whether to deviate the train and kill one person or maintain the direction and kill five persons on the track*").

[38] O. Boissier, et al. "A roadmap towards ethical autonomous agents". (2015): available at https://ethicaa.greyc.fr/media/files/ethicaa.delivrable.3.pdf.

Last, the technological approach understandably discusses AI and robots with foreseeable business and consumer application. As a result, the discussion concentrates on thin *classes* of AI applications. This has a major downside. It gives moderate consideration to open-ended AIs with multiple functionalities, like IBM's Watson.[39] More fundamentally, the findings reached under this approach may be exposed to rapid obsolescence, if AIs follow a process of technological convergence. This would happen if AI applications morphed, combined or converged: for instance, one can easily envision a proximate world where speech recognition will be used in conjunction with exoskeletons and virtual reality devices. A more distant projection is one in which one's digital driver is also one's digital butler and partner.

C. COMPARISON

The issue of whether AIs and robotic applications deserve to be granted legal rights – for instance, to litigate, contract or own property – and/or duties helps picture the differences between the legalistic and technological approach.[40] Under the former approach, the social planner would start from rules that confer legal personhood on humans, corporations, international organisations and innate objects like trees,[41] and attempt to understand if similar legal status can be extended to AIs and robots. Under the latter approach, the social planner looks at the technology, and applies a Turing test to establish whether legal personhood can be granted. This approach was the one followed by Lawrence Solum when he considered the following thought experiment: "*Could an artificial intelligence serve as a trustee?*".[42]

The implications of both approaches are clear. The legalistic approach is driven by *teleological* questions: courts and legislature are asked to consider the goals pursued by legal personhood law.[43] For instance, legal personhood was granted to corporations in order to promote economic exchange. A question that will therefore arise will be: shall we give AIs and robots legal personhood to promote economic exchange, as was done for corporations?

---

[39] *See*, for instance, http://www.rossintelligence.com/.

[40] S. Chopra and L. F. White, *A legal theory for autonomous artificial agents*. University of Michigan Press, 2011 at 155 (conferring legal personhood necessitates a "*decision to grant an entity a bundle of rights and concomitant obligations*").

[41] C. D. Stone, *Should Trees Have Standing? Law, Morality, and the Environment*, 3rd Ed, Oxford University Press, 2010.

[42] L. B. Solum, "Legal Personhood for Artificial Intelligences", *N.C. L. Rev.* 70 (1992): 1231-1288.

[43] *See* Chopra & White, *supra* at 186 ("*the decision to accord or refuse legal personality (both dependent and, in function of increasing competence, independent) would ultimately be a result-oriented one for courts and legislatures alike, and cannot rest solely on conceptual claims*").

Similarly, a certain degree of legal personhood has been recognized to trees on grounds of sustainable development. In turn, the social planner could ponder whether AI legal personhood is likely to contribute to the conservation of global resources.

In contrast, the technology approach is more *ontological*. The inquiry consists in an assessment of what the technology is, and in particular of whether the technology displays *human* features.[44] The discussion thus mostly consists in a reflection on our own selves, and what makes us human.[45] With this, the technology approach may be skewed towards anthropomorphic AI and robotic applications ("*hard*" robots like Asimo) or representations (softbots like Siri, Alexa or Cortana), when it discusses the pros and cons of conferring of legal personhood to AIs.

C. SUMMATION

The legalistic and technology approaches are not exclusive and can certainly be combined. I represent them here as polar cases to ease understanding of the main methodological options available to a social planner.

Often, however, regulators seem to follow both. This can be seen in the European Parliament 2017 resolution on Civil Law Rules on Robotics.[46] Even though the methodology followed remains unknown, the rich inventory of measures proposed in relation to AIs and robotics seems underpinned by both sets of perspectives.

II. REGULATORY TRADE-OFFS

Law and regulation in a context of technological emergence involves complex trade-offs. I discuss each of them first from a theoretical perspective, and follow with an application to AIs and robots.

---

[44] Here, we fall in a complex philosophical discussion, as to whether our criteria of choice is Wittgenstein "family resemblance" or Aristoteles theory of definition (known as predicables doctrine), which focuses on genus and specific essence. *See* Ayers, Michael R. "Locke versus Aristotle on natural kinds." *The Journal of Philosophy* 78.5 (1981): 247-272.

[45] This issue has been thoroughly discussed by early philosophers since Aristoteles up until enlightenment. Those debates consist in a reflection of whether humans and animals are different by a matter of degree (as suggested by Darwin's theory of evolution, or are distinct in kind. Many properties have been underlined to denote the specificity of the human king: thought, language, instinct, self-consciousness, emotions, perfectibility, religion, vertical position, etc. For a good review (and rebuttal), see Philalethes. "The Distinction between Man and Animals." *Anthropological Review* (1864): 153-163.

[46] *Supra* European Parliament resolution of 16 February 2017.

A. DISABLING REGULATION

Regulation can nip technological progress in the bud. Several variants of this phenomenon are documented in the literature. A historically well-known example is the Red Flag Act of 1865 which sought to regulate the introduction of automobile in society.[47] The Statute imposed a 2 mph speed limit on locomotives in urban agglomerations and required a pedestrian carry a red flag 60 yards in front of the vehicle. It is often said that this regulation stifled the development of the British motor car industry for as long as 30 years.

Closer to us, Renda and Pelkmans take the example of Genetically Modified Organs ("GMOs").[48] In the EU, a strict authorization regime was set. As a result, only two new GMO products have to date been allowed to be cultivated, despite reported benefits to farmers and a possible reduction in poverty and hunger.[49]

Even closer to digital innovations, the regulation of bitcoin payments and blockchain platforms, including caps on usage and transparency requirements may actually run amok with the very purpose of anonymity sought by users of the system.

With this background, let us try to consider possible examples of AI-disabling regulations. Some of them are already in discussion. Restrictions on the ability or conditions for the operation of unmanned aerial systems and drones to fly beyond visual line of sight ("BVLOS") could stifle the development of delivery and transport applications. Similarly, requirements to allocate a given amount of money to a robotic application could prohibitively inflate the cost of production and consumption of AI applications, and raise a barrier to their penetration in society.

An interesting observation to be made is that regulation can be consciously or even deliberately disabling. To put the point differently, regulation is not only inadvertently or accidentally disabling. GMOs are a case in point. One of the rationales for their strict regulation is that the scientific community suspects a safety threat, primarily of the

---

[47] D.G. Victor, T.C. Heller, and N.M. Victor, "Political Economy and the Hydrogen Revolution" Working Paper 17, 2003, available at http://pesd.fsi.stanford.edu/sites/default/files/pol_econ_H2.pdf; N-G. Rodrigo. "No Bad Deed Goes Unrewarded: Cause, Consequence, and Deviance in Emerging Technological Regimes." In *Questioning Causality: Scientific Explorations of Cause and Consequence Across Social Contexts* (2016): 343-360.

[48] J. Pelkmans, and A. Renda. "Does EU regulation hinder or stimulate innovation?." *CEPS Special Report* 96 (2014).

[49] Those are the NK603 GM maize and the Amflora potato.

environment and of human and animal health.[50] This rings interesting parallels with the safety risks of certain robotic applications like LAWs.

B. KNEE-JERK REGULATION

Knee-jerk regulation consists in adopting inefficient over-regulation in response to risks, incidents and accidents.[51] The Fukushima nuclear disaster is a case in point. It combines a classic "*dread risk*" (radioactivity), a punctuating event (the disaster itself), and a resultant stigmatization (essentially, by the public opinion).[52] In several western countries, communities have blamed nuclear technology and not Japan's exposition to seismic hazard.[53] As a result, some countries relatively immune to seismic and weather events have introduced nuclear exit policies, and turned to fossil fuel energies as an alternative.

Knee-jerk regulation can also occur in the face of anticipated risks, and this sometimes blurs the distinction with disabling regulation. Consider reproductive cloning. The fear that cloning can alter the course of humanity through interference with reproduction – together with ethical and religious concerns – has fuelled calls for strict regulation.[54] In 2005, the General Assembly of the United Nations adopted a Declaration on Human Cloning, which called upon Member States to adopt all measures necessary to prohibit all forms of human cloning inasmuch as they are incompatible with human dignity and the protection of human life.[55] In the same spirit, the European institutions have sought to prohibit the issuance of any patent rights related to reproductive cloning.[56]

---

[50] M. Lee, *EU regulation of GMOs: Law and decision making for a new technology*. Edward Elgar Publishing, 2009, at 79.

[51] Also known as the "*risk regulation reflex*". Interestingly, knee jerk responses work also in the other direction. President Obama mocked the Republicans calls for deregulation as a knee jerk obsession: "*Feel a cold coming on? Take two tax cuts, roll back some regulations and call us in the morning*". *See* https://www.whitehouse.gov/the-press-office/2012/09/07/remarks-president-democratic-national-convention

[52] E. Hammond, "Nuclear Power, Risk, and Retroactivity", *Vanderbilt Journal of Transnational Law* 48: 1059-1082.

[53] W. J. Kinsella, W. J. "Being "Post-Fukushima": Divergent Understandings of Sociotechnical Risk." Fukushima Global Communication Programme Working Paper Series (2015).

[54] P. Sevanthinathan, "Heavy Regulation of Human Cloning as an Alternative to a Complete Ban." *Quinnipiac Health LJ* 10 (2006): 219.

[55] *See* General Assembly adopts United Nations Declaration on Human Cloning by Vote of 84-34-37, GA/10333, 8 March 2005.

[56] E. Jackson, *Regulating reproduction: Law, technology and autonomy*. Bloomsbury Publishing, 2001.

The potential for knee jerk regulation of AIs and robotics is easy to foresee. Take a deficient AI airliner autopilot. And let one assume that society displays a lower tolerance threshold for accidents caused by machines. In this context, it can be anticipated that society will respond to any crash with a prohibition of fully or partly AI-operated planes, and roll back to require a significant degree of human operation. This is in spite of existing evidence that human operated flights may be significantly less secure than MI-assisted ones, and that the source of the problem often lies in the complex interaction between automated machines and humans.[57] With this, instead of prohibiting, regulation should seek to improve machine-human cooperation in ways that enhance safety.

C. RENT-SEEKING AND REGULATORY CAPTURE

The rent seeking hypothesis explains that private interest groups have incentives and ability to steer government regulatory action towards their own benefit. Rent seeking is effective when the private interest group is a large organization, for the benefits of regulation will exclusively accrue to it. In contrast, when private interests are fragmented, rent seeking is less likely because the costs of coordination are high, and the benefits of regulation must be shared.[58]

Frederic Bastiat, a XIXth century intellectual, described rent seeking with a powerful allegory.[59] In a "*Petition from candle makers*", the French lightning industry was seeking protection from the ruinous competition of the sun… The candle makers had accordingly petitioned the Chamber of Deputies to "*pass a law requiring the closing of all windows, dormers, skylights, inside and outside shutters, curtains, casements, bull's-eyes, deadlights, and blinds—in short, all openings, holes, chinks, and fissures through which the light of the sun is wont to enter houses, to the detriment of the fair industries with which, we are proud to say, we have endowed the country* [...]".[60]

In the modern world, examples of rent seeking are all around us: the car dealer industry has sought enforcement of protective legislation against Tesla's B2C model; taxi drivers' have

---

[57] D.A. Mindell, *Our robots, ourselves: Robotics and the myths of autonomy*. Viking Adult, 2015.

[58] G. Stigler, "The Theory of Economic Regulation", *The Bell Journal of Economics and Management Science*, Vol. 2, No. 1 (Spring, 1971), 3-21.

[59] *See* Library of Economics and Liberty, Economic Sophisms, at http://www.econlib.org/library/Bastiat/basSoph3.html.

[60] In the candle makers' views, this competition was allegedly stirred by England, clouded under the fog ("*This rival, which is none other than the sun, is waging war on us so mercilessly that we suspect he is being stirred up against us by perfidious Albion*").

requested additional protection from the competition of ride-sharing apps like Uber and Lyft. And the hospitality sector has sought to steer municipalities across the world to undermine the operation of services like RBnB or of platforms like Booking.com and Expedia.[61]

In so far as AIs and robots are concerned, rent seeking concerns seem more acute under the technological than the legalistic approach. In general, the technological approach seems more skewed in favour of technological development, in contrast with the legalistic approach which seems more technology neutral. Hence, the technological approach may be more exposed to rent seeking by stakeholders from the technology community.

Concretely, one area with rent seeking potential is car insurance. In many countries, the law imposes insurance duties on driver and/or user. With self-driving cars, the case for driver and/or user compulsory insurance is less compelling. There is less driver control, fewer accidents and lower damages at society level.[62] Of course, trees and snow still fall, causing casualties on the road. However, as autonomy progresses, allocating liability on driver and/or user seems less justified, and a transfer to driverless cars manufacturers is a more plausible

[61] The development of the e-cigarette is a case in point. This technology has an undisputable disruptive effect on the tobacco industry. *See* B. Yandle and R.E. Meiners, and J.H. Adler, and A. P. Morriss, "Bootleggers, Baptists, and E-Cigarettes" (January 1, 2015). Case Legal Studies Research Paper No. 2015-3. Available at SSRN: http://ssrn.com/abstract=2557691. In this context, there is little wonder why firms like Trierenberg-Gruppe, the world's top producer of cigarette papers, and other trade associations representing manufacturers and retailers the tobacco industry have expended resources to lobby for tougher EU rules. *See* http://corporateeurope.org/sites/default/files/attachments/tobacco_lobbyists_all_fired_up_ahead_of_key_vote.pdf. In this context, there is little wonder why firms like Trierenberg-Gruppe, the world's top producer of cigarette papers, and other trade associations representing manufacturers and retailers the tobacco industry have expended resources to lobby for tougher EU rules. What is more unexpected, however, is that manufacturers with smoking cessation products joined the fray, despite their alleged commitment to public health. For instance, it is reported GlaxoSmithKline lobbied for tougher EU rules on the grounds that e-cigarettes encouraged young people to start smoking. The result is an EU regulation that effectively authorizes Member States to prohibit the "*placing on the market of specific electronic cigarettes or refill containers*". *See* Directive 2014/40/EU: Article 20(2) of 2014 tobacco directive: "*Manufacturers and importers of electronic cigarettes and refill containers shall submit a notification to the competent authorities of the Member States of any such products which they intend to place on the market*"; Article 20(11) of 2014 tobacco directive: possibility for MS authorities to take "*appropriate measures*" if "*serious risk to human health*"; Recital 46, appropriate measures include "*the prohibition of the placing on the market of specific electronic cigarettes or refill containers*".

[62] All the more so if the law incrementally reduces individuals' freedom to drive. *See* D. McCrum, Insurers will destroy themselves to nudge us into robot utopia, Alphaville, 4 March 2014, at https://ftalphaville.ft.com/2014/03/04/1787962/insurers-will-destroy-themselves-to-nudge-us-into-robot-utopia/.

option. Moreover, the insolvency concern that underpins the *compulsory* nature of insurance seems less problematic with car manufacturers.[63] The problem, of course, is that insurance companies have much to lose if compulsory driver and/or user insurance is abandoned. Their relative bargaining power against a handful of manufacturing companies is much lower than in relation to myriad individual drivers and/or users.[64] Last, car manufacturers exposed to hold-up conduct by insurance companies, may have incentives to vertically integrate into insurance services, rendering the insurance industry irrelevant in the long term. This situation incentivizes insurance companies to lobby in favour of an extension of compulsory driver and/or user insurance for self-driving cars.[65]

D. REGULATORY TIMING

An often encountered critique of regulation is it inability to keep pace with technological evolution. The point is: regulation is a reactive process. When adopted, it may already be obsolete.[66]

More generally, the regulatory process often struggles to keep abreast of technology. Mandell explains that emerging technologies present a "*quandary*": whilst a benevolent social planner would like to discourage research in harmful technologies and incentivize research in beneficial ones, the risks and opportunities created by emerging technologies cannot be "*suitably understood until the technology further develops*".[67] The regulatory process must therefore keep a degree of "*connection*", and wait for technology to develop so as to endow the social planner with enough knowledge.[68] However, as the social planner acquires the

---

[63] M.G. Faure, "Economic criteria for compulsory insurance." *The Geneva Papers on Risk and Insurance Issues and Practice* 31.1 (2006): 149-168.

[64] *See* https://www.euractiv.com/section/digital/news/eu-considers-new-insurance-laws-for-driverless-cars/.

[65] By the same token, car manufacturers may lobby government so as to be insulated from all liability, and deflect it towards software developers.

[66] Consider a proposed Federal Aviation Administration ("FAA") regulation that insists that unmanned aircraft system ("UAS") must be limited "*to daylight-only operations, confined areas of operation, and visual-line-of-sight operations*". Now compare this with surveys reporting that pilots spend just seven minutes manually operating their planes in a typical flight. *See* http://www.nytimes.com/2015/04/07/science/planes-without-pilots.html?_r=2. No wonder why companies like Amazon, Intel and Google have railed against emerging drones delivery regulation, which they consider outdated. *See* http://www.theverge.com/2016/1/16/10777144/delivery-drones-regulations-safety-faa-autonomous-flight.

[67] G.N. Mandel, "Regulating emerging technologies." *Law, Innovation and Technology* 1.1 (2009): 75-92.

[68] R. Brownsword and M. Goodwin, *Law and the Technologies of the Twenty-First Century. Texts and Materials*, Cambridge University Press, 2012.

necessary knowledge, the technology entrenches and it may be too late to act. This problem is known as the Collingridge paradox.[69]

This risk is one discussed by Nick Bostrom as the "*treacherous turn*". This notion refers to the pivot point which is reached when a recursive self-improving AI becomes sufficiently strong to strike humans without warning or provocation.[70] In a matter of minutes, a malignant AI may consider that humans are threats to the achievement of its final values and turn against them avoiding the controls systems set by engineers. Bostrom uses the example of an AI designed to optimize production in a paperclip factory. Following a treacherous turn, the AI would proceed by first "*converting the Earth and then increasingly large chunks of the observable universe into paperclips*".[71]

E. ENABLING REGULATION

Even free-market spirited scholars claim that the absence of regulation can thwart technological evolution. There are several narratives of this claim. For some, regulation is a necessity, because the alternative – a form of case by case litigation – is worse.[72] At bottom, the point is that litigation prospects disincentivize research into potentially beneficial technology. Product liability litigation in relation to deficient medical devices is an often heard worry.

Others argue that the application of default rules to emerging technologies is unclear, in particular at early phases of technological development. In turn, this legal uncertainty allegedly has an adverse effect on investments.[73]

A last school insists on the potential of regulation to actually enable innovation. A popular theory is the so-called Porter hypothesis which states that strict environmental, health and safety standards prompt firms to improve their productivity, and finds that "*properly designed*

[69] D. Collingridge. *The Social Control of Technology*. London: Francis Pinter Ltd., 1980. Biotechnologies in the large sense are an example. Whilst they promise many wonders in public health, nutrition and environmental protection, the creation of new eco-systems could precipitate the demise of animal or human species, create new diseases, etc.

[70] N. Bostrom, *supra* at 144-145.

[71] *Id*. at 150.

[72] Epstein for instance poses the necessity of regulation: **"***At bottom, the proper inquiry never poses the stark choice of regulation versus no regulation***".** *See* R.A. Epstein *supra*.

[73] *See* R. Calo, 2011, *supra*: "*legal uncertainty could discourage the flow of capital into robotics or otherwise narrow robot functionality*".

[regulatory] standards can trigger innovation that may partially or more than fully offset the costs of complying with them".[74] In Porter's view, "*tough standards trigger innovation and upgrading*", and prompt firms to re-engineer. In addition, strict regulatory standards can promote market competition, by inducing firms to race for first movers' advantages.[75]

In so far as AIs and robots are concerned, Ryan Calo has warned of the risk of crippling legal liability regimes.[76] Uncertain liability rules could act as disincentives to investment into open robotics markets, and channel the flow of capital towards *narrow* robot functionality where producers can better manage risk, leaving open robotics underdeveloped. Calo militates in favor of a specific immunity regime for AI and robotics manufacturers, close to the immunities enjoyed by firearms producers and website operators.

III. LIABILITY

It is now time to take a practical look at the challenges faced by a social planner in a context of AI and robotic emergence. To keep the discussion concrete, I have chosen to discuss the question of liability, which is the legal issue that garners the highest interest from scholars and practitioners.[77] In short, the question of liability consists in asking: who pays for harm caused by AIs? In this section, I tackle this simple interrogation first by considering the main regulatory options faced by a benevolent social planner operating under the basic legal structure (A). I then discuss first order (B) and second order problems (C).

A. BASIC LEGAL STRUCTURE

In both civil and common law systems, the law imputes liability for harm upon a variety of legal agents. Put simply, victims can seek damages against several targets. *First*, liability can

[74] M.E. Porter, and C. Van der Linde. "Toward a new conception of the environment-competitiveness relationship." *The journal of economic perspectives* 9.4 (1995): 97-118.

[75] N.A. Ashford, and R. P. Hall. "The importance of regulation-induced innovation for sustainable development." *Sustainability* 3.1 (2011): 270-292. Pelkmans and Renda document empirical examples of enabling regulation. One of them is the regulation of end-of-life vehicles. Under the EU regulation, ambitious recycling targets were adopted far in excess of industry anticipations, including the reuse and recycling of 85% of cars by 2015. As a result, automotive manufacturers engaged in a virtuous cycle of innovation at design and planning stage. *See* J. Pelkmans, and A. Renda, *supra*. The optimistic tone of the literature on enabling regulation shall however not obscure that firms may follow innovation strategies designed to evade the law. The 2015 Volkwagen NOx (nitrogen oxides) emission scandal highlights that when overly ambitious regulatory targets are adopted, firms have incentives to invest into technologies which game the enforcement system, including malicious software.

[76] R. Calo, 2011, *supra* .

[77] This is true both under the legalistic and technological approaches.

be imputed on the *perpetrator* of harm. Under civil liability rules or the law of tort, one is liable for damages caused by his or her own acts. A certain degree of wrongdoing is necessary to trigger the perpetrator's liability, be it *negligence* (breach of duty of care or omission), *fault* or a *breach of a statutory duty*.[78] Transposed to the subject of this paper, this means obtaining damages from the AI or robot itself. To date, however, seeking a judicial declaration of liability against an AI or a robot is a dead-end. This would require the preliminary recognition of legal personhood to AIs and robots, a bold step that no legal system has yet undertaken.

Second, a victim can seek to engage the vicarious liability of a *third party* with oversight. For simplicity, I call this third party a "*governor*". Common examples include the vicarious liability of employers, parents, masters and owners, for damage caused respectively by employees, children, slaves and property. In so far as AIs are concerned, the governor of an algorithmic machine could be held liable for damages. In both civil and common law, vicarious liability regimes are negligence or fault-based. They necessitate a degree of wrongdoing. For instance, in Belgian civil law, when the agent of harm is a *thing*, the damage claimant must establish a "*defect*" understood as an abnormal feature that may be conducive to damage in certain circumstances.[79] In contrast, in French civil law, there is no necessity to establish a defect. Both regimes, however, require proof that the keeper of the thing exercised effective *control* over it.

Third, victims can target damages claims against the *manufacturer*. In most countries, defective products laws repute manufacturers liable for damage caused by the products they bring to markets.[80] A product is generally deemed defective when it generates unexpected injury in normal use.[81] In turn, a defect is assessed in standard circumstances of use that is "*use to which it could reasonably be expected that the product would be put*". Most defective products laws establish a *strict liability* regime: the mere proof of a defect triggers liability,

[78] For civil law, I rely on Article 1382, 1383 and 1384 of the Belgian and French civil codes.

[79] Hof van Cassatie, A.R. nr. C.95.0250.F, *Lacroix t. Gemeenschappelijk Motorwaarborgfonds*, 1 maart 1996.

[80] *See*, in the EU, Directive 85/374/EEC on liability for defective products.

[81] EU law says that a defect means that the product "*does not provide the safety which a person is entitled to expect*".

regardless of whether there has been fault or negligence.[82] And defective product laws tolerate only a limited number of exoneration causes.[83]

The basic legal structure thus essentially offers victims two potential and non-alternative routes for the recovery of damages caused by AIs and robots. Victims may seek to impute liability to the *governor* of the application and/or to its *manufacturer*. Having come to that initial diagnosis, a social planner may reach the preliminary conclusion that AI and robotic liability is not a legal desert. The next exercise that it would then face is one of problem identification. This is the set of issues that I now examine, distinguishing between first and second order problems.

B. FIRST ORDER PROBLEMS

First order problems are those that arise when legal rules and standards fail to fulfil their goals. In other words, first order problems are teleological.

Generally, liability law can be deemed to serve two functions: a *corrective* function, namely remove past harm by providing a solvent target to victims; an *incentive* function, namely deter the future occurrence of damage, by confiscating the gains of harmful conduct.[84] Against this background, the basic legal structure generates three types of first order problems when applied to AIs and robots.

The first problem is one of under-correction of harm. Assume that open, generalist and modular AI applications (hereafter, "open AIs") are introduced in society, besides narrow AIs like self-driving cars, delivery drones or intelligent vacuum cleaners. Open AIs lack a *set function*. Like Swiss army knives, they can be used for many purposes, in many contexts, and in combination with many other tools. They therefore cannot be deemed to operate abnormally or unexpectedly, as requested to trigger liability under torts, civil law of defective products legislation. In such circumstances, victims will fail to collect compensation, and one goal of the liability system is put into jeopardy.

---

[82] *Id*. "*liability without fault on the part of the producer is the sole means of adequately solving the problem, peculiar to our age of increasing technicality, of a fair apportionment of the risks inherent in modern technological production*".

[83] In EU law, the manufacturer can argue in defense that it did not put the product into circulation, that the defect has been caused by mandatory compliance with rule of law, or that scientific or technical state of the art was not such as to enable the existence of the defect to be discovered.

[84] *See* generally G. Williams, « The Aims of the Law of Tort », 4 Current Legal Problems, 1951

Relatedly another under-correction problem could occur with fully autonomous AIs that do not implicate a human "*in*" or "*on*" the loop. In such circumstances, damage claimants will face a daunting task to establish a situation of *governance* under vicarious liability. By definition, a fully autonomous AI or robot is one that evades human control. Here, one can see instantly that such problems do not necessarily require clarification, modification or introduction of new liability laws. Instead, safety standards and product design requirements are a possible avenue. For instance, such statutes could require all fully autonomous AIs and robots to be equipped with fail-safe, red button shutdown systems.

The second problem is one of over-correction of harm. As a mirror hypothesis to the previous problem, courts may consider that the user of a generalist AI may reasonably expect safety in all circumstances, which risks being impossible to guarantee by robotic producers. This is particularly true of open robotic platforms which invite third parties to tinker. As Calo puts it: "*the manufacturer could not necessarily anticipate the universe of potential problems that might stem from third party innovation and provide warnings or modify the platform design in response*".[85]

The third problem is one of perverse incentives in a complex value chain. The Coase theorem helps understand this.[86] Pursuant to the Coase theorem, *who* is liable under the law – victim, governor or manufacturer – is to some extent irrelevant. As long as there are no transaction costs (and property rights are well defined), parties will bargain the efficient solution for society. This can be understood with the following fictional example: A is a crop grower who uses a robot gardener to irrigate its fields at night, when water and electricity costs less. B is an entrepreneur who buys a neighbouring piece of land, and launches a boutique hotel. Because of A's robot's noise and lights at night, B's clients flee in droves and write incendiary reviews on websites.

Coase explains that the question of who should be liable in that situation is a moral issue that should not retain the interest of a social planner. This is because, as long as transactions costs are minimized, A and B will reach agreement over the cheapest solution regardless of the allocation of liability. Again, the example is useful. To eliminate the harm, two options are possible: A sends the robot for mechanical soundproof improvement, and this will cost €5,000; B installs double glazing, and this will cost €15,000. In our example, the *socially*

[85] *See* R. Calo (2011) *supra*.

[86] R. Coase, "The Problem of Social Cost", *J. Law & Econ*. 3 (1960): 1-44.

efficient solution – the one to which the lower amount of resources are devoted – is that the robot is sent for mechanical improvement. Coase shows that this solution will prevail, whatever the law says of the allocation of liability. If the law assigns a right to silence to B, then A will not wait to be sued,[87] and he will pay 5,000€ to retool its robot. Now if the law assigns a right to noise to A, then B will pay €5,000€ to A in exchange for a mechanical improvement of the robot.[88] In either case, the efficient solution prevails, regardless of who is liable. For the social planner, the specific allocation of liability is therefore irrelevant.

However, there is a significant caveat to this logic. Negotiation is not always frictionless. If there are transactions costs (which is often the case in the real world), parties will not be able to bargain efficient outcomes, and the social planner should allocate liability upon the *cheapest cost avoider*. Returning to the example, transactions costs would for instance be present if A was a cooperative of distinct crop growers. Being liable under the law, B would find bargaining for €5,000 with multiple parties costly. In turn, B would likely contemplate installing double glazing, which would lead to a €10,000 loss for society. In such circumstances, Coase said that the law should assign liability so as to emulate the cheapest cost solution that would have prevailed in negotiation.[89] In our example, A should be deemed liable.

If one applies this idea to AIs, the Coase theorem invites a social planner to consider whether (i) there will be transaction costs to negotiation; and (ii) if this is the case, to search for the cheapest cost solution, and more particularly, to look into the AI and robotic value chain for liability on the cheapest cost avoider. Whilst the first issue is largely an empirical matter, the second question is easier to conceptualize. A variety of economic agents other than those that the basic legal structure reputes liable – the *governor* and *manufacturer* – contribute upstream to the operation of AIs upon the world. This is the case of programmers (C), standard setting organizations (D), and perhaps the AI application itself (E). And those agents may be in a position to avoid the harm at lower cost. For instance, the cheapest solution in the example may consist in C writing a line of code that disables all noisy robot functionalities at night or for D to write a standard to the same effect.

[87] For €15,000.

[88] In reality, B will pay an amount comprised between 5,000€ and 15,000€.

[89] One underpinning of this idea is that the negative externality inflicted by regulatory system should be as minimal (in efficiency terms) as possible.

This latter point also illustrates that the solution to an apparent liability issue does not need be an adjustment of the *law*, but may consist in promoting specific robot ethics solutions, at the design stage.[90]

C. SECOND ORDER PROBLEMS

Second order problems are implementation frictions that arise before courts until first order issues are resolved by the social planner. Consider the following example: during day to day operation, A's robot gardener acquires private data (pictures essentially) over B's clients. A court may be faced with the question whether A's robot gardener is a data *"processor*" according to the rules on privacy, which can in turn be held liable to compensate any damage due to unlawful data processing.[91] Or say that A's robot gardener has shot sunset photographs during its operations. A displays the pictures on its website. B copies the pictures, and uses them to advertise its boutique hotel on reservation platforms. A sues B for copyright liability. B counter argues that A has no right to damages, because it is the robot gardener that is the "*author*" entitled to copyright protection.

In the absence of a clear, unequivocal and definitive resolution of first order issues by a central lawmaking authority – Supreme Court, deliberative assembly, executive agency – such questions are left to litigation. Courts will solve implementation problems on the basis of preferred principles of legal interpretation (*eg* ontological, deontological, consequential or teleological). Note here that second order problems may arise before first order problems.

The main problem generated by second order problems is one of legal uncertainty. Pending the resolution of goal-related issues by a central lawmaking authority, a degree of variance is inevitable in decentralized courts system. This may raise compliance costs. As seen above, a

[90] All the more so if, on top of efficiency concerns, the social planner seeks to tackle morality issues.

[91] In EU law, a processor can be defined as: "*processor' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*". *See* Article 4(8) of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

popular claim is that legal uncertainty is nefarious to investments,[92] and that regulation is therefore needed to allay those pitfalls.[93]

This claim, however, must be put into perspective with the benefits arising from the decentralized operation of the courts' system. One of them is the supply of empirical data to lawmakers, in areas rife with information asymmetries, and thus exposed to misguided regulatory intervention.[94] Another one is the production of scientific evidence, thanks to the examination and cross-examination of facts, law and witnesses under due process rules (meaning that arguments get a chance to be falsified). A last one is the generation (and experimentation) of alternative options for lawmakers.

Given its mixed effects, legal uncertainty does not in and of itself constitute a sufficient basis to warrant regulatory intervention.[95] In practice, this means that lawyers' concerns at the repetition of second order implementation problems ought not trigger immediate response. A period of experimentation in the courts system may be an appropriate alternative. All the more so if this creates time for the parallel deliberation of complex normative issues.

IV. FRAMEWORK FOR REGULATION AND NORMATIVE APPLICATIONS

[92] Even free-market scholars consider that absent regulation, risk of case by case litigation can be worse. *See* R.A. Epstein, *supra* ("*At bottom, the proper inquiry never poses the stark choice of regulation versus no regulation*"); J.M. Balkin, *supra* ("*The new technology disrupts the existing scene of regulation, leading various actors to scramble over how the technology will and should be used. As people scramble and contend with each other over the technology, they innovate—not only technologically, but also socially, economically, and legally—leading to new problems for law. Instead of saying that law is responding to essential features of new technology, it might be better to say that social struggles over the use of new technology are being inserted into existing features of law, disrupting expectations about how to categorize situations*").
Economists also explain that unpredictable regulatory intervention is a source of uncertainty, which too can be detrimental to investments.

[93] In substance, they request the central lawmaking authority to specify *ex ante* a precise legal *rule* – the robot is a processor if it meets condition X, Y and Z – or an abstract legal *standard* – the robot is a processor if it is *autonomous*.

[94] They do this because legislators generally have limited information and expertise in specialist areas and, though knowing that there is a problem, tend not to know how to solve it.

[95] In line with this, regulation theorists rarely mention legal uncertainty as a market failure worthy of Government remediation.

This section attempts to provide some guidance on a possible method for the regulation of AIs and Robots. I introduce some key concepts first (A). I then delineate some preliminary normative implications (B).

A. CONCEPTS

To lay down my framework, I start from the conservative but not exhaustive proposition that law and regulation purport to address externalities. By externalities, I mean activities that inflict harm or provide benefits to third parties.

This proposition is rooted in mainstream public interest theory. The choice of this framework is not the result of convenience or coincidence, but instead follows the dominant paradigm in AI and robotics.[96] Sci-fi readers will recall that Asimov's first law of robotics ambitioned to prevent a robot to "*injure* a human being, or, through inaction, allow a human being to come to *harm*". Since then, much scholarship in "*hard*" and "*soft*" social sciences discusses Asimov's laws as a starting point of inquiry. Rarely, in fact, has pop culture so much influenced academic research.

Two types of externalities can be distinguished. A negative externality occurs when an AI or robotic application imposes costs on third parties, and when the AI or robotic application (or its governor) do not internalize all or any of those adverse effects. A positive externality appears when an AI or robotic application provides benefits on third parties, and when the AI or robotic application (or its governor) fail to appropriate all or any of those effects. Economic theory suggests that rational agents overinvest in the supply of activities that produce negative externalities. For example, competing manufacturers may race to introduce new generations of robots at rapid pace, in disregard of the costs incurred by users to decommission obsolete robots. Conversely, economic theory indicates that rational agents underinvest activities which yield positive externalities. For example, manufacturers may not invest in ethical standards and "*friendly AI*" initiatives for robotic applications, because the benefits of this are largely appropriated by third parties. In both configurations, economic theory explains that a public interest-driven social planner can attempt to correct externalities through the imposition of taxes, the allocation of subsidies or the promulgation of explicit legislative and administrative controls.[97]

---

[96] *See* D. Weld, and O.Etzioni. "The first law of robotics (a call to arms)." *AAAI*. Vol. 94. No. 1994. 1994: 1042-1047.

[97] R. A. Posner, "Theories of Economic Regulation", *Bell J Econ* 5 (1974), 335-358.

Building on this notion of externalities, I introduce hereafter a novel distinction between three groups of externalities. The first group consists in *discrete externalities* that is harm or benefits with the following non-cumulative properties: personal, random, rare or endurable. By personal, I consider externalities that affect the unit level of the *individual*. Randomness means that the externality may affect all and any third party with *equal chance*. By *rare*, I consider low frequency externalities. And by *endurable*, I talk of externalities that do not completely "*ruin quality of life*" or that do not radically improve it (non pure human enhancement).[98]

A typical example of a negative discrete externality is a robot gardener whose visual recognition module dysfunctions, and confuses the neighbor's cat with a parasite, ending up spraying the cat with toxic pesticide. A typical example of a discrete positive externality occurs if the crop-grower is able to eradicate nocturnal parasites when the robot gardener is operated at night.

The second group covers *systemic externalities.* This refers to third party harm or benefit with the following non-cumulative properties: local, predictable, frequent or unsustainable. By local, I look at harm or benefit that affect a non trivial segment of the population. By predictable, I envision harm or benefit that is foreseeable for a benevolent social planner. By frequent, I mean a repeated occurrence of harm or benefit. By unsustainable, I refer to a non-transitory reduction or increase in well-being of the local population class under consideration (given scarce resources). A durable rise in inequalities (poor get poorer, rich get richer) is a case in point.

An often-discussed negative systemic externality consists in the substitution of man by intelligent machines on the factory floor (and the ensuing disappearance of many existing manufacturing jobs, pressure on workers' wages in the long term, etc.). Conversely, a less discussed, though equally important positive systemic externality consists in the new complementary jobs that will be created by the introduction of intelligent machines and cognitive computing in industrial sectors (and the corollary reduction in manufacturing costs across the economy as well as transfers of productivity gains to consumers through lower prices).

The third group of externalities comprises existential threats and opportunities created by AIs and robotic applications. To denote their existential nature, I call them as "*existernalities*".

[98] N. Bostrom, "Existential risk prevention as global priority." *Global Policy* 4.1 (2013): 15-31.

Existernalities have several cumulative properties: they are global, improbable, unpredictable and terminal. By global, I mean that existernalities hit indiscriminately across geographies, ethnical communities and social organizations. By improbable, I refer to the idea that *rational* wisdom often dismisses existential AI and robotic risks as fictional. By unpredictable, I have in mind that we fail to assess the timescale and likelihood of existernalities. By terminal, I envision the potential of existernalities to extinguish humanity as we know it.

Negative existernalities include the risk of human extinction,[99] malign superintelligences,[100] robot killers and warriors, and other dystopian, terminator-spirited scenarios of machine takeover. Positive existernalities include pure human enhancement,[101] cosmic endowment,[102] virtual immortality, etc. Often, the boundary between a positive and negative existernality is a subjective issue. For instance, time-travel is seen by some as a threat for humanity, and by others as an improvement.

In the table below, I list some examples of discrete and systemic externalities, as well as existernalities.

[99] *See* R. Kurzweil, "The Singularity is Near", 2005. This also covers dehumanization through the blurring of distinctions between machines and humans.

[100] *See* N. Bostrom, 2013, *supra*.

[101] Pure human enhancement goes beyond the restoration of destroyed human functions. Human enhancement is often opposed to therapy, which "*aims to fix something that has gone wrong*". But this distinction is not airtight. *See* N. Bostrom, and R. Roache. "Ethical issues in human enhancement." *New waves in applied ethics* (2008): 120-152.

[102] *See* N. Bostrom, 2013, *supra*.

Table 1: Typology and Examples of Externalities

Discrete Externality		**Public interest**
Negative	An industrial robot restarts abruptly and kills a worker on the factory floor.	
Positive	Drone spots thief on way to delivery destination, alerts law enforcement which stops the burglar.	
Systemic Externality		
Negative	General reduction in privacy across society due to generalized operation of information-hungry AI systems	
Positive	Improved disaster responses and humanitarian systems thanks to AI monitoring of population w/o consent	
Existernality		**Existential**
Negative	Permanent state of war following introduction of Lethal Autonomous Weapons ("LAWs")	
Positive	Acceleration towards technology frontiers: time-travelling; emulated minds; cosmic exploration	

B. NORMATIVE IMPLICATIONS

In this last section, I develop some normative implications of the above conceptual framework. This is done by taking the perspective of a public-interest minded social planner.

To start, the resolution of discrete externalities should be left to the basic legal infrastructure. In practice, the social planner defers to the decentralized courts system which will process discrete externalities on a case-by-case basis. Disputes are solved *ex post* through the application of the general rules of property, contract and liability and other specific laws. This is acceptable because discrete externalities cannot affect society by any significant order of magnitude. Moreover, this regulatory approach is efficient, because it allows a degree of decisional experimentation, benchmarking, and cross-fertilization.

In contrast, the more severe threshold effects encountered with systemic externalities deserve a degree of *ex ante* consideration by the social planner. The question before it is whether *ad hoc* law or regulation ought to be adopted to correct the systemic externality. Here are some examples of such questions in relation to negative externalities: must a specific tax be

introduced in robotic-intensive industries subject to creative destruction?; must black-box requirements be imposed on manufacturers of robots confronted with moral dilemmas like the trolley problem?; must specific privacy regulation be adopted on the second-hand robot market to protect data subjects, including previous governors? And here are more examples for positive systemic externalities: given the public goods nature of infrastructure and collective action problems amongst competing producers, must subsidies be allocated for the construction of controlled environments for robots (for example, specific road infrastructure for driverless cars)?; should developers and manufacturers of generative AI technologies enjoy statutory immunity for damages caused by their inventions?[103] Should intellectual property regimes be relaxed, to enable open, transparent and peer-scrutinized research processes in AI applications, in furtherance of the goal of friendly AI?

Regulatory responses to systemic externalities must be subject to *ex ante* and *ex post* impact assessment. By *ex ante* impact assessment, I refer to the prospective cost-benefit evaluation of future regulatory options. By *ex post* impact assessment, I consider retrospective cost-benefit measurement of experimented regulatory options. In this variant, the social planner experiments the various regulatory options in dedicated zones of the real life environment, and proceeds to evaluate the results of such tests. In Japan, for instance, the creation of so-called "*Tokku zones*" system has entitled robot manufacturers to conduct practical tests on public roads and environments.[104] This mixed *ex ante* and *ex post* approach limits risks of Collingridge type dilemmas and reduces the potential of disabling regulation.

Existernalities create concerns of such levels that they can be *ex ante* subject to law and regulation, without prior AI and robotic experimentation, implementation or realization. Given their *global* nature, the regulation of existernalities should tentatively be decided by international organisations ("IOs"). However, IOs are often characterized by gridlock on existential issues (like peacekeeping or climate change) due to their wide membership. In the AI field, endless discussions have taken place at the United Nations over a proposed ban on the use of lethal autonomous weapons ("LAWs"). Regional consensus-building institutions (like the EU) might thus be better forums for the initial regulation of existernalities.

[103] For a precedent, *see* Senate Bill: S. 1458 (103rd) General Aviation Revitalization Act of 1994, known as GARA.

[104] Y-H. Weng, et al. "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 (2015): 841-857.

In addition, the fact that existernalities are "*black swans*" implies that democratic deliberation institutions which are the transmission belt of mainstream opinion currents may fail to anticipate them.[105] Conversely, knee-jerk regulatory responses cannot be excluded in such participatory systems. A *degree* of expert and technocratic input in decision-making seems therefore appropriate. In this context, the involvement of standard setting organisations (like the IEEE, SAE, the ISO and many others) may play a useful contributive role to the definition of early positions on existernalities. Last, objections to *ex ante* intervention are not material, because the costs of type II errors (false negatives) in relation to existernalities are higher than the costs of type I errors (false positives). A type II error occurs when the social planner fails to remedy a serious existential risk in probability and/or intensity terms. A type I error occurs when the social planner wrongly remedies a moot existential risk in probability and/or intensity terms. Immediately one understands that the cost of a type II error is existential, whilst this is not necessarily the case for a type I error. The cost of the latter is thus more acceptable than the cost of any type II error which will always be existential. But there is more. A type II error in relation to existernalities is not reversible, because humanity has disappeared. This excuses any and every type I error in relation to existernalities.

V. CONCLUSION

This paper has attempted to sketch a rudimentary framework for the law and regulation of AI and robots. Its ambition is primarily pragmatic: help decision-makers assailed with requests for intervention in this emerging field understand *if* and *how* regulation is needed.

The proposed framework and its tentative implications are based on a public interest approach. I propose to index the intensity of regulatory response upon the nature of the externality created by an AI application. When AI-generated externalities are *discrete*, social planners should defer to *ex post* litigation before courts. When AI-generated externalities are systemic, social planners should envision *ex ante* regulation, but carefully test and experiment. This meshes the benefits of anticipation and empiricism, and avoid Collingridge dilemma as well as disabling regulation problems. Last, when AI-generated externalities are *existential*, social planners should seriously consider *ex ante* intervention, and bring into it a degree of expert deliberation.

[105] Events that come as surprises to most, if not all. *See* N. N. Taleb, *The black swan: The impact of the highly improbable*. Vol. 2. Random house, 2007.

In future research, I expect to sophisticate this framework through case-studies in various legal systems and disciplines within legal systems (*eg*, tax, consumer protection, intellectual property, standardization and other fields). I also hope to further illustrate and sophisticate the various declinations of externalities explained in this paper. Last, I will attempt to give a less simplistic picture of the choice faced by the social planner, and in particular move beyond the schematic representation of a trade-off between the basic legal structure on the one hand and full-blown regulatory intervention on the other hand. This will necessitate to bring on board bargaining institutions, private ordering and many other empirical realities to which this paper did not (could not) do much justice.

*

* *