Closing Pandora’s box? The EU proposal on the regulation of robots

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published in:
Pandora's Box - The journal of the Justice and the Law Society of the University of Queensland

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Closing Pandora’s box? The EU proposal on the regulation of robots

BURKHARD SCHAFER

I OF ROBOTS, MYTHOLOGY AND THE LAW

“Whereas from Mary Shelley’s Frankenstein’s Monster to the classical myth of Pygmalion, through the story of Prague’s Golem to the robot of Karel Čapek, who coined the word, people have fantasised about the possibility of building intelligent machines, more often than not androids with human features [...]”

Thus starts the Motion for a European Parliament Resolution with recommendations to the Commission on Civil Law Rules on Robotics, submitted by the Committee on Legal Affairs. (henceforth: “the Motion”).1 This paper will attempt a first analysis of key notions of the proposal, by following the proposers in exploring the emerging discourse on robot regulation through the prism of literature and mythology.

The unusual and somewhat tongue-in-cheek introduction is an appropriate reminder of just how much our thinking about robots and their legal regulation is influenced by their depiction in mythology, literature and film. For centuries, we have projected our hopes and fears into human-like machines, seeing in the back-reflection from their metallic (typically) faces an account also of what we are or as what we see ourselves. Law and legal regulation plays a consistent theme in these stories, as we will see below.

II CLOSING PANDORA’S BOX

A reference missing from the Motion is that to the story of Pandora. When juxtaposed to the Genesis account of Eve and the Fall, we get a first idea of the concerns that the committee tries to address. Pandora was created by Hephaestus, blacksmith to the gods and master-engineer. She was designed with one purpose in mind – punishing mankind for acquiring the fire from the gods. Built to allure and to seduce, her task is to deliver the jar that contains all the evils in the world, "burdensome toil and sickness that brings death to men". Only hope remains in the jar before she closes it again.

The story told by Herodotus bears some striking resemblances, but also a crucial difference to that of the Biblical Eve. Eve too is designed – though maybe better described as bioengineered. She too will bring through her actions toil, sickness and disease into the world. But unlike Pandora,

---

who acted strictly according to her instructions, with Eve it is the ability for autonomous decision making and with that the ability to act in ways unforeseeable to her creator that causes the harm.

Today we find the same topoi in the highly gendered depiction of robots in film - Ava, the robot in Ex Machina, just like Pandora is designed to seduce, just like Eve ultimately through her autonomy bringing doom on the naïve man she interacts with. Maybe even more worrying, real life examples of robotics follow the same patterns, with the female Siri and Tay playing the role of secretary, while the male chatbot Ross\(^2\) delivers proper legal advice.

Exploring how gender and fear shape our perception of robots would go beyond the scope of this paper.\(^3\) But Pandora and Eve, each in their own way, encapsulate the fears and concerns that dominate the legal debate on robotics: One is the concern that malicious designers could develop deadly robots intentionally to inflict harm on humans. Concerns that have led to calls to outlaw military application of robotics. The other is the fear that by creating entities with autonomy and permitting them to act in ways that may be in principle unpredictable by us, we are not only engaging in risky behavior, we might sever the nexus between creator and creation that allows us to attribute legal liability and responsibility if things go wrong. Just as Eve's action plays a central role in Christian apologetics, so robot autonomy is seen as a potential “get out of jail card” that could be played by manufacturers or sellers if their products cause harm.

Or, in the words of the committee on the reasons for taking action at sec. 24:

> “whereas, notwithstanding the scope of the Directive 85/374/EEC, the current legal framework would not be sufficient to cover the damage caused by the new generation of robots, insofar as they can be equipped with adaptive and learning abilities entailing a certain degree of unpredictability in their behaviour, since these robots would autonomously learn from their own, variable experience and interact with their environment in a unique and unforeseeable manner.”\(^4\)

The fear is that this could expose buyers and the general public to harm without recourse. It could equally however create uncertainty for manufacturers, sellers and investors that prevents the robotics industry realising its beneficial potential. Law, appropriately adjusted to this new reality, might be able to give us reasonable hope in a safe robotic future. Yet hope, as Pandora's story shows, is an ambivalent concept. It is unclear if by keeping hope in the jar, Pandora denied us “even hope” and punished us even more, or if closing the lid was an act of kindness – after all, hope was placed by Zeus in a jar that contained all evils, and what is more evil than hope continuously disappointed. Is the Motion then aimed to “close the box” in the sense of keeping a lid on potential harm? Or is it giving us “false hope”, in that it deludes us into thinking that by regulating a technology we can make it safe?

Or is the role of the proposal merely symbolic, a form of “red flag” law that does not address any real problem, but responds to public concerns by creating unnecessary and to a degree burdensome duties on robotics manufacturers, but with the advantage of increasing public acceptance of the technology? These are some of the themes that this paper will explore.

---


\(^3\) See, eg, Friederike Eyssel and Frank Hegel,‘(S) he's Got the Look: Gender Stereotyping of Robots’ (2012) 1 *Journal of Applied Social Psychology* 42(9), 2213-2230.

\(^4\) *Civil Law Rules on Robotics Motion*, 6.
III THE LAW AND THE GOLEM

The EU Motion prominently mentions the story of the Golem as one of the oldest examples of man-made autonomous machines. If the Golem was the first robot, then a lawyer was the first roboticist. Tractate Sanhedrin 65b from the Talmud (the “cases and materials” of Jewish law) describes how amora (legal scholar) Rava created in the 3rd century BCE a person-like being from mud. Rava was one of the most influential law teachers of his time, contributing to the canon of Talmudic law a concept of good faith acquisition of lost/abandoned property; and to the laws of civil procedure a secularised notion of witness credibility assessment. That it should be a lawyer who created the first robot is within the Jewish religious framework entirely understandable: God is (perfect) law and the ultimate creator, so everyone who achieves near-mastery of the law could arguably also achieve near-mastery of the art of creation. Near-mastery only though, and indeed Rava’s robot quickly failed the Turing test: When Rava, maybe to put his work to the test, sends him on an errand to another influential amora, Rav Zeira, the man-machine is quickly found out. Incapable of answering questions directed at it, Zeira easily identifies the originator behind the ploy: "You were created by the sages; return to your dust". Then and now, the ability to display human characteristics when under cross-examination and responding appropriately to questions was the litmus test that distinguishes man from machine; then and now, mastering language turns out to be a difficult task to achieve.

The idea of lawyers as arch-roboticists quickly disappears from history, though it returns briefly in the 19th century with an interesting twist. Legal formalism developed an ideal of the judge as adjudicator that saw them as machine-like in nature, working through simple algorithms, ideally available in codified form, to determine the right outcome without fear, favor, or any other emotion for that matter. Roscoe Pound dismissively termed this “mechanical jurisprudence”, but for some of its adherents this epithet would have been a source of pride rather than disparagement. Over 100 years before Pound, Julian de la Mettrie wrote in “Machine Man”:

“To be a machine and to feel, to think and to be able to distinguish right from wrong, like blue from yellow [...]”

Rather than making robots like their Jewish predecessors, the legal formalists of 18th Europe dreamt of turning themselves into machines. From this two themes emerged that are also relevant for the contemporary discussion on robot regulation.

The first is the idea, central for formalist jurisprudence, that legal codes can be seen as a library of rules, which together with an appropriate logic form an algorithm that can determine mechanically the outcome of a case. This idea, which informed the development of first generation legal expert systems such as Taxman or the Latent Damage System, also opened up the possibility of a different approach to robot regulation. Rather than using law only retrospectively, after a violation has occurred, implementing formal representations of relevant legislation in the robot’s software might ensure law compliance by design. This idea was popularized in literature through Asimov’s famous Laws of Robotics – though we should note that their main narrative function is to create problems and to require workarounds. Asimov did not advocate them as a solution, if anything, his stories show how difficult it can be to reduce normative decision making to simple rule

---

7 Julien Offray de La Mettrie, Man Machine and Other Writings (Cambridge University Press, 1996) 35.
following. However, the idea got traction in the legal domain. The first commercially deployed example was through DRM as a form of copyright law by design, and more recently through the “privacy by design” requirement, encouraged by implication in the EU Data Protection Directive, and soon to be explicitly mandated in Art 25 of the EU General Data Protection regulation. Lessig’s influential (and highly critical) appraisal of software-enforced rule compliance finally brought the equivalence between legal and software code into the mainstream discussion on technology regulation.

The Motion to the EU Commission remains deeply ambivalent on this issue. Citing explicitly Asimov, it states as General Principle I that

“whereas, until such time, if ever, that robots become or are made self-aware, Asimov’s Laws must be regarded as being directed at the designers, producers and operators of robots, since those laws cannot be converted into machine code.”

It is unclear why the committee thinks that self-awareness is a precondition for legal rule following. It is true that the rules in the form given to them by Asimov are too general and abstract to be suitable candidates for a formal capture that could guide machine behaviour. However this does not mean that quite sophisticated formal representations of legal norms can’t under the right conditions be an effective tool for robot regulation. The committee seems to recognise this when at sec.10, it also

“calls, in this regard, on the Commission to foster the development of standards for the concepts of privacy by design and privacy by default, informed consent and encryption;”

It seems clear that despite the dismissive reference to Asimov’s laws in the general part, some form of legal reasoning capacity will have to play a role in the tool set for efficient robot regulation. Given just how prevalent robotic devices are bound to become, we will therefore likely face a future of “ambient law”, where gadgets, cars, automated homes or smart cities constantly run algorithms that are isomorphic formal representations of relevant legal provisions.

The second lesson that we can learn from the 19th century idea of turning lawyers into robots concerns the effect of robotics on the labour market. We find the idea that modern working practices are turning humans into machines prominently in Karel Čapek’s play R.U.R. from 1920, which the EU Motion also cites. The heroine of the book, Helena, is a representative of the “League of Humanity”, a human rights organisation that lobbies for employment and other civil rights for robots, including the right to get paid fair wages. The impact of these robots on the labour market and wider society is however as profound as it is ambivalent, resulting in deskilled humans with decreasing birth rates and ultimately little to protect them when the robot uprising begins. While for Čapek, robots were a metaphor for dehumanising working conditions under modern modes of production, the fear that robots will disrupt our labour markets and put additional strain on already overstretched social security systems also plays a central role in the current debate.

“Bearing in mind the effects that the development and deployment of robotics and AI might have on employment and, consequently, on the viability of the social security systems of the Member States, consideration should be given to the possible need to introduce corporate reporting requirements on the extent and proportion of the contribution of robotics and AI to the economic results of a company for the purpose


11 Civil Law Rules on Robotics Motion, L.

of taxation and social security contributions; takes the view that in the light of the possible effects on the labour market of robotics and AI a general basic income should be seriously considered, and invites all Member States to do so.13

With this the Motion opens up for discussion two of the more radical proposals for a wider societal response to increased automation at the workplace. One is to abandon the notion of employment as the norm, and of wages as the typical form of income. Instead, a general basic income is suggested as an alternative should, as some commentators have predicted, the reduction in available jobs overwhelm existing social security networks.14 A possible source for funding of such a scheme is also hinted at by the EU committee. Robots could be treated as employees for tax and social security purposes. While not exactly getting paid, as Helena lobbied for in R.U.R, robots could be paying income tax – or rather a levy could be raised from companies that splits the difference between the reduction of costs that the company gains through automation and the costs that this creates for the welfare system through loss in tax revenue and increased demand for unemployment benefits.

This suggestion raises an interesting philosophical question with direct legal relevance: How do we count robots, and how do we identify individual specimens? If a company owns one hundred cars, each with identical software, all communicating constantly with each other and a central server, is this one (distributed) robot, or one hundred? If the latter, why would an autonomous car, which will have several hardware and software components that constantly talk to each other and a central processor, not also count as several robots? Furthermore, as the software of a robot will in most cases require constant updating, since software ages faster than hardware. But is a robot that undergoes a radical change in its software still the same – or should it be considered as a new employee?

Čapek’s story also brought to the fore the possibility that robots should be recognised as legal persons, and with that another important crossover between law, literature and robotics. Using the law to resolve conflicts caused by the autonomy and intelligence of machines is a recurrent theme in 20th century robot literature. It is through a legal trial that Commander Data in the Starrek universe has to prove that he is deserving of legal protection and the status of a legal person. Formal confirmation of citizenship and the rights and duties that it entails to the robot Johnny Five brings the “Short Circuit” franchise to a conclusion. These and similar stories evidence how much we still trust the law as a vehicle to settle social and political disputes. In a case of life imitating art, the mayor of Nanto City granted in 2010 the therapeutic seal robot Paro a “koseki” (household registry/birth certificate), which lists Shibata Takanori, Paro’s inventor, as the robot’s father.15

The idea of robots as holders of rights has been mooted on and off in the academic discussion for quite some time, but never attracting significant support.16 A Horizon scanning report for the UK government however took the idea serious enough to contemplate limited civil rights for robots within the next 50 years.17 At first sight the EU Motion appears to follow this line of thought and asks to at least contemplate the possibility of

---

13 Civil Law Rules on Robotics Motion, 23.
14 On basic income and robot technology see in particular, James Hughes, ‘A strategic opening for a basic income guarantee in the global crisis being created by AI, Robots, desktop manufacturing and BioMedicine’ (2014) 1 Journal of Evolution and Technology 45.
15 For a discussion see Jennifer Robertson, ‘Human rights vs. robot rights: Forecasts from Japan’ (2014) 46 Critical Asian Studies 571.
“creating a specific legal status for robots, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons with specific rights and obligations, including that of making good any damage they may cause, and applying electronic personality to cases where robots make smart autonomous decisions or otherwise interact with third parties independently.” 18

Together with the notion of robots as tax payers, this idea led commentators in the popular press to the conclusion that the EU is indeed preparing the ground for legal recognition of robots, possibly in analogy to the legal status of corporations. A more cautious reading of the proposal however would replace “right” in the above section with “capacity”, in particular the capacity to enter into legal agreements that are binding on the owner. As a simple example, we can think of an automated car that pays any applicable road tax “on its owner’s behalf”.

If read like this, we can see that the discussion is far from new. At the turn of the century, advances in the design of autonomous agent software led to concerns about the legal status of contracts that were negotiated between machines, with no or limited human oversight. The “law of electronic agents” workshop series that was held as part of the EU funded Agentlink network addressed these issues comprehensively. The emerging consensus at the time indicated that radical solutions such as ascribing legal personality to software agents was unnecessary, and that existing legal instruments were capable of dealing with machine-to-machine contract negotiations in an equitable way.19

IV ROBOTS OR AI

Why would the committee feel the need to reopen this discussion? Part of the reason is the very definition of “robot” that the Motion suggests, and to which we will turn our attention now.

The Motion asks the Commission to

“propose a common European definition of smart autonomous robots and their subcategories by taking into consideration the following characteristics of a smart robot:

- acquires autonomy through sensors and/or by exchanging data with its environment (inter-connectivity) and trades and analyses data
- is self-learning (optional criterion)
- has a physical support
- adapts its behaviours and actions to its environment”

If adopted, the EU would be the first jurisdiction with a generic definition of robot, to be applied across legal domains. While a small number of jurisdictions has defined the term “robot” in law for specific purposes, they typically are to be found in highly technical laws that deal with issues such as their treatment for tariff purposes (Russia, which uses a rather long and cumbersome definition), facilitate and encourage investment in robotic technology or set aside physical spaces where they can be safely tested. None of these definitions play a direct role in core civil law, or are intended to regulate liability of/for robotic devices. The EU proposal also comes close to definitions used by roboticists. Mataric for instance defines robots as “an autonomous system

18 Civil Law Rules on Robotics Motion,31F.
which exists in the physical world, can sense its environment, and can act on it to achieve some goals.”

While the proposed definition is in line with that used within the technology community, reasonably flexible to anticipate future developments and sufficiently precise, it is nonetheless questionable if it is adequate for legal purposes. It is premised on the idea that there are certain legally problematic aspects of robots that apply across all or most applications. But are there really legal questions that military robots, care robots for the elderly, medical robots, automated cars, toy robots and advanced washing machines share? The main concern of the committee is clarifying civil liability, but even for such a limited objective, it seems obvious that very different rules apply to different types of robots, or indeed to the same type of robot used in different scenarios and by different actors. Some, but not all robots will simply be consumer goods, and significant parts of their liability hence regulated by consumer protection law. Others will be used by law enforcement and military, including dual-use robots, where the liability regime in many countries creates special liability rules and exemptions. Medical devices and cars traditionally have their own regulatory regimes that already implement some of the suggestions that the committee proposes.

Two key components of the committee’s recommendations are a mandatory insurance scheme. Manufacturers or owners insure the robots against harm to the buyer and to third parties. In addition, a supplementary fund is suggested to cover those machines for which no insurance was taken out. In what is possibly the most innovative suggestion in the proposal, monies paid to robots as part of their work could directly flow into this fund, and cover it automatically in case it causes actionable damages. This idea of legally protected funds that “follow the robot” seems to be inspired by the Roman law of slavery. The peculium in Roman times was a fund slaves (or indeed sons) could be given to manage for themselves. While ultimately, they were still part of the property of the paterfamilias, in practice, they functioned like property of the slave and could be used to buy his freedom. The committee is not suggesting this use of the peculium, nor do they seem to envisage a peculium that contains again other robots (the way the Roman peculium could). Instead, we could imagine maybe an entry into a blockchain ledger, which then would allow claimants with little bureaucratic efforts or cost to claim “directly” against the robot, without having to determine if the fault was due to software or hardware, the liability the seller’s or manufacturer’s.

The second component is a mandatory registration scheme, for all robots, for the purpose of “...ensuring that the link between a robot and its fund would be made visible by an individual registration number appearing in a specific EU register, which would allow anyone interacting with the robot to be informed about the nature of the”

This too could be facilitated through blockchain ledger technology. The problem with this proposal, and one that the committee tacitly admits when it asks the Commission to develop an appropriate classification scheme to determine which machines should be subject to such an approach, is that for those robots that will pose the greatest risks – medical robots and cars – registration and insurance systems already exist. On the other hand, a requirement to register and insure every individual Romba, washing machine or robotic toy dog seems vastly excessive.

While the suggested general definition is therefore overly inclusive, and in need to be broken up again by the suggested classification scheme, it is in another respect overly exclusive. “Unembodied AIs”, intelligent software agents, are not covered by the definition. Siri or Tay the Apple and Microsoft chatbots, are (probably, but see below) not covered by the definition. This is problematic not only because AIs like these will play such a significant role in changing the way we interact with

---

21 Committee Motion 31 a and b.
23 *Civil Law Rules on Robotics Motion,31E*. 
Information Technology. It also ignores that some of the most pressing legal issues that the Motion tries to address are not only the same for disembodied AI, but have been analysed, discussed and in some cases actioned on successfully in this field. As we saw, this is particularly the case for the question of legal personhood. This issue arises always when an entity, be it machine or software code, is not only to a degree autonomous, but also has the ability to communicate. By excluding unembodied AI, the committee forgoes the chance to learn from the experience legal systems have made with regulating software agents. The most important of these is maybe that the status of the entity itself is less relevant for the discussion. What matters is the status of the speech acts they perform. Once we have decided we want to “count as” contractual offers machine generated speech, the issue of the status of the machine becomes almost irrelevant and we have the choice to treat it as a mere message, as an exercise of the law of agency or indeed “directly” to a legal person. Similarly, if a machine utters the words “I hereby declare you man and wife”, it will be a question for administrative or canon law to decide if this can count as a valid performance of a wedding. No general definition of what a robot is in law can or should pre-empt this discussion, to which different legal systems and traditions may well give different answers.

Excluding disembodied AI from the remit of the discussion not only prevents the committee to learn from experience made with these devices, it also prevents a discussion of aspects of the law that may be in much more need of revision than contractual or even core delictual liability.

When Tay, the Microsoft chatbot, was released on Twitter, it quickly picked up (or rather, was forced to pick up through a concerted effort by some users) particularly loathsome ideas and habits. This also means that some of its Tweets could be speech acts of a different legally relevant kind: defamation or criminal insult, or in jurisdictions with relevant legislation criminal hate speech or Holocaust denial. Nobody suggested suing Microsoft though – too obvious was the fact that “Tay was still learning”. In the UK, defamation is a strict liability tort however, and to bring the legal ideal in line with the practical reality of learning machines, one might consider carving out a “court jester” exemption to robot speech in those cases at least when it is clear that the utterance was a result of imperfect learning.

Whenever a robot produces speech, this also creates questions for copyright law. The EU Motion mentions IP but briefly and concludes that in all likelihood, robotics technology is not creating significantly new problems. This is a surprising omission, also given the importance of IP to stimulating technological investment. Some robots produce works of art, for instance Taida, winner of the 2016 Robotart competition. So far, only the UK, and soon Japan, provide for explicit rules on the status of creative works generated by autonomous machines, and whether or not their approach is helping or hindering the evolution of the field needs discussion. Even more important than works of art, the overwhelming amount of data is generated automatically, by autonomous devices and through machine-to-machine communication. Some of this data has commercial value. If my driving trains the AI in my automated car, and the manufacturer can get access to this data to improve their products, do I have a proprietary interest in this data if it is not personal information and protected by data protection law? The commission Motion focuses on questions of liability, but even here IP law matters. For robots will not just be producers, they will also be “consumers” of IP protected work. They are reliant on input from their environment to navigate, learn and improve. Some of this information will in turn be IP protected. Can a drone take images of famous buildings for navigation purposes and share copies with its fleet, potentially violating the IP rights of the architect? Can they data-mine my tweets to improve their speech?

With these observations, our discussion turns full circle. The first robot, the Golem, lacked capacity to communicate. This also meant that many legal issues were pre-empted. The Motion for an EU initiative on robot regulation is an important step to open up the discussion on appropriate, harmonised responses to the robot revolution. Many of its ideas are bold and worthy of discussion, even if few are likely to make it into law. Yet at some crucial points, a major reassessment is

necessary. Neither the proposed definition of robot, nor the subsequent focus on liability, seem to have identified some of the most intricate problems or the most pressing legal needs. To start a discussion on legal regulation of robotics with references to robots in literature and mythology was an unusual step to take. It has significant pedagogical advantages, as it reminds us of the fears and hopes that mankind projects into its machines, fears and hopes that then put pressure on legislators to act. Yet, more might have been learned by taking these stories more seriously, with their focus on man-machine cooperation and with that robotic speech. Of greater concern maybe is however that they also create in our mind a vision of robot that for the foreseeable future will be the exception rather than the norm: anthropomorphic, with high degrees of autonomy, multi-purpose and competing rather than cooperating with humans. For this type of robot, the committee proposal makes bold and innovative suggestions worth of further exploration by the EU Commission. But we should be concerned that by extending this regulatory approach across the whole range of robots, we could impede needlessly innovation and investment in some fields, or conversely, this over-extension of the proposal could undermine its merit in those fields where more radical legislative action is beneficial and needed.