Winter December 28, 2018

# Robot Law 1.0: On Social System Design for Artificial Intelligence

Yueh-Hsuan Weng, *Tohoku University*

Yueh-Hsuan Weng (2018) Robot Law 1.0: On Social System Design for Artificial Intelligence, Woodrow Barfield, Ugo Pagallo (Eds.), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Publishing, ISBN 978-1-78643-904-8

# Robot Law 1.0: On Social System Design for Artificial Intelligence

Yueh-Hsuan Weng
FRIS, Tohoku University

## Introduction

The focus of this chapter is on the issue of how artificial intelligence (AI) may influence the evolution of a legal framework for human-robot co-existence. In this chapter I use three legal paradigms, Robot Law 1.0, 2.0 and 3.0, to analyze law and policy as related to AI in different technical stages of development. As an important issue, under the category of Robot Law 1.0, I discuss the concept of "Social System Design," a strategy that aims to improve integrated governance for embodied AI. As artificially intelligent robots become more common in human society, it will be especially important to consider the ethical, legal, and social impacts implicated by the design of such systems. The concept of Social System Design will not only benefit legislators and policy makers, but also lead to an efficient regulation of intelligent robots and their AI-driven risks. As another laudable goal, Social System Design will also be helpful towards embedding social values into increasingly intelligent robotic systems. Thus, given this design paradigm which is based on a human-centered value system, artificially intelligent robots will more likely abide by moral obligations, an important goal for artificial intelligence researchers and society in general. Issues of morality accompanies the Robot Sociability Problem, which as discussed in this chapter has many facets, and suggests that an interdisciplinary approach towards thinking about the design of emerging intelligent sociable machines is necessary.

## 1. Background

Artificially intelligent systems manifest in many forms, and many would agree that increasingly intelligent robots are an important application of artificial intelligence; this observation motivates the interest of legal scholars who advocate for a law of artificial intelligence. In the mid-20th century, the term "robot law" began as a movement to develop moral principles for robots, such as to obey their master when making their own decisions while performing tasks in daily scenarios in the service of humans. One well-known paradigm for robot law is Asimov's Three Laws of Robotics – an ethical guideline that teaches intelligent robots how to ensure the safety of human-robot interactions based on a human-centered point of view. However, Asimov's three laws have changed greatly during the last three quarters of a century, that is, since Asimov first introduced the Three Laws of Robotics in his classic work "Runaround" in 1942[1]. Based on Asimov's robot laws and the work of scholars, many ethical branches began to grow at the root of

[1] Isaac Asimov. Runaround. Astounding Science Fiction, 1942. Reprinted in Isaac Asimov. I, Robot. Ballantine Books, New York, 1983.

his three laws – not the least of which was AI safety – an important discipline within robot ethics. The IEEE Global Initiative for Ethical Considerations in AI/AS is an example of government regulations which provide diverse social values in terms of "Ethically Aligned Design"[2]. Additionally, over time, the meaning of "robot law" began to change from machine ethics for intelligent robots into a broader definition and categorization, including laws and rules for design, manufacturing, usage, the interaction of intelligent robotic technologies with other smart technologies, and additional ethical considerations for other forms of artificial intelligence.

Other legal issues of concern for increasingly intelligent robots and other smart technologies have also been the subject of discussion among legal scholars. For example, professor of law Ryan Calo described the correlation between the openness of a robotics platform and its corresponding tort liability given an accident[3]; Ugo Pagallo discussed the concept of privacy risks for robotics technology[4], Susanne Beck provided an interesting perspective of criminal laws for robotics[5], Woodrow Barfield pointed out ethical and legal concerns for Cyborg technology[6], van den Hoven van Genderen discussed the ancient Roman slaves system and how it relates to legal personhood rights for robots[7], and Erica Fraser and Burkhard Schafer described how to judge the IP innovation rights implicated when an AI entity is involved[8]. Following this trend towards identifying legal problems for AI and robotic technologies, the author's interest developed on how artificial intelligence influences the evolution of legal systems, which is the focus of this chapter

## 2. The Pyramid of Robot Intelligence

From a neurologists' point of view, the human brain has three layers -- primitive, paleopallium, and neopallium -- that operate like "*three interconnected biological computers, [each] with its own special intelligence, its own subjectivity, its own sense of time and space, and its own memory*".[9] Based on this view, an analysis of the hierarchical taxonomy of robot intelligence and how these factors can influence robot law in a long-term perspective is of interest to this chapter.

---

[2] The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems, Version 1, IEEE, 2016. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

[3] Ryan Calo, Open Robotics, Maryland Law Review, vol.70, no.3, 2011

[4] Ugo Pagallo, Robots in the cloud with privacy: A new threat to data protection? Computer Law & Security Review, 29 (5), 501-508, 2013

[5] Susanne Beck, Intelligent agents and criminal law—Negligence, diffusion of liability and electronic personhood, Robotics and Autonomous Systems, vol.86, 138-143, 2016

[6] Woodrow Barfield, Cyber-Humans: Our Future with Machines, Springer, 2015

[7] Robert van den Hoven van Genderen, Robot as a Legal Entity, Legal Dream or Nightmare? Proceedings of the 20th International Legal Informatics Symposium – IRIS 2017, page 161-170, February 23-25, University of Salzburg

[8] Erica Fraser and Burkhard Schafer, Self-made (Machine) Men – IP Implications of Inventions by Robots, Proceedings of the 20th International Legal Informatics Symposium – IRIS 2017, page 171-178, February 23-25, University of Salzburg

[9] Paul D. MacLean, The triune brain in evolution: role in paleocerebral functions, Springer, London, 1990

"Action Intelligence" is located at the bottom of the robot intelligence pyramid. Its functions are analogous to human nervous system responses that coordinate sensory and behavioral information, thereby giving a robot the ability to control body movement, move spatially, operate its arms to manipulate objects, and visually inspect its environment. The next, or second, level of Autonomous Intelligence refers to capabilities for solving problems and involves pattern recognition, logical reasoning, machine learning, and planning based on prior experience. "Intelligence" for robots (and other smart technologies) can also be referred to as "weak AI" or "narrow AI" in which a robot performs specific tasks in a narrow domain but with remarkable abilities. Examples include Google DeepMind's AlphaGo[10] and IBM's super computer Watson[11].

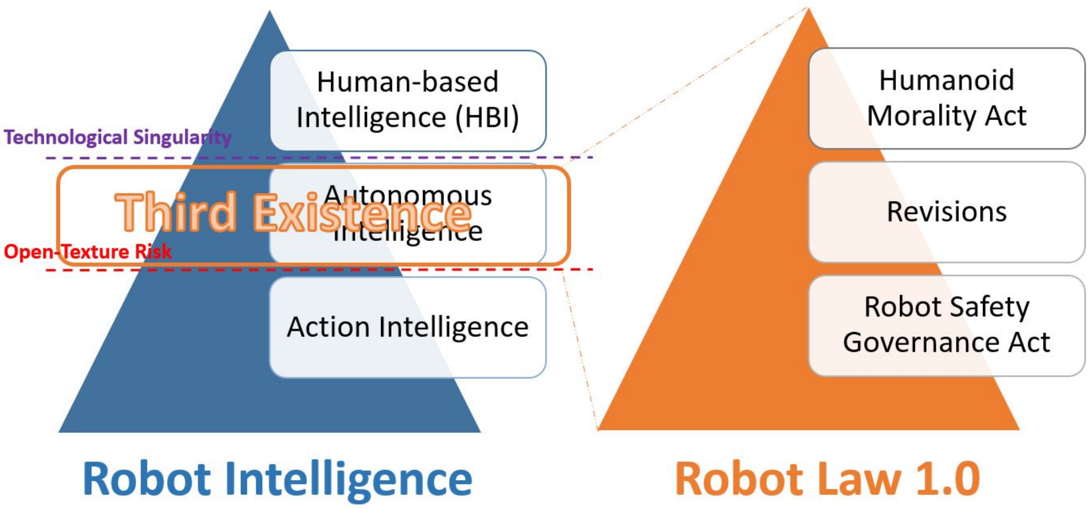


**Figure 1. Two Pyramids for the Governance of Embodied AI.**

Referring to Figure 1, robots with Action or Autonomous Intelligence are neither self-aware nor have their own internal value system in which to guide them in deciding what is right or wrong. However, a distinct difference between robots with Action versus Autonomous Intelligence is based on their adaptiveness to unstructured environments. The latter are superior in performing their tasks without predefined information about the real world, such as human living spaces. This difference leads to autonomous robots' having adaptive behaviors and to become "open-textured", and leads to the possibility of many safety and ethical hazards in the real world, and also forms a boundary known as Open-Texture Risk[12] which refers to risks from robots' autonomous and potentially harmful behaviors. Such complex, changeable and unpredictable behaviors cause the gap shown for modern machine safety in risk assessment. The word "open-texture" originally

[10] AlphaGO, DeepMind, https://deepmind.com/research/alphago/
[11] Watson, IBM, https://www.ibm.com/watson/
[12] Yueh-Hsuan Weng, Chien-Hsun Chen, Chuen-Tsai Sun, The Legal Crisis of Next Generation Robots: On Safety Intelligence. In: Proceedings of the 11th International Conference on Artificial Intelligence and Law (ICAIL'07), Stanford, CA, pp 205–209

came from the field of law, and refers to the ambiguity of interpretation of legal terminologies[13]. Specifically, it is difficult to give direct and explicit definitions to some legal clauses due to the openness and evolving nature of language (especially when technology is involved). The same principle applies to the behavior of "Autonomous Intelligence" robots' as well. Their dynamic decision-making abilities leads to difficulty when trying to specify potential risks possible from all of their behaviors – this topic is what the current ISO 12100 standard focuses on, that is, machine safety[14]. In other words, Open-Texture Risk is different for machines which can perform adaptive behaviors within unstructured environments and its range can be expanded by increasing its intelligence level.

Referring to the pyramid shown in Figure 1 for the governance of embodied artificial intelligence, at the third level of robot intelligence is Human-Based Intelligence (HBI) which refers to higher cognitive abilities that will allow robots creative ways to look at their environment and also the ability for abstract thought, even leading to consciousness. This level of robot intelligence is referred to in the literature as "strong AI" or "Artificial General Intelligence (AGI)". With advancements in artificial intelligence, the manifestation of strong AI in the form of Superintelligence or "Artificial Super Intelligence (ASI)" may be possible. Based on futurist Nick Bostrom's definition, Superintelligence is "*any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest*"[15].

In general, the Technological Singularity is a hypothesis that relates to the growth and evolution of technology ultimately causing an intelligence explosion which would then lead to a powerful form of Superintelligence that is far beyond human intelligence.[16] However, different definitions have been proposed for the Technological Singularity[17]. For example, according to Vernor Vinge and Ray Kurzweil's definition, the Technological Singularity is more closely linked to ASI[18] In contrast, artificial intelligence expert Ben Goertzel describes The Singularity by focusing on the potential for an uprising of AGI[19]. Hence, depending on the definition, the Technological Singularity may include both possibilities of either the birth of human-level AGI or superhuman ASI- only time will tell. Murray Shanahan argued that once human-level intelligence has been achieved, the transition to Artificial Super Intelligence could be very rapid[20]. In my view, from a regulation perspective, the coming of human-level AGI is a more important development

[13] David Lyons, Open texture and the Possibility of Legal Interpretation, Law and Philosophy, 18(3), 297-309, 1999
[14] Gurvinder S. Virk, S. Moon, R Gelin, ISO Standards for Service Robots, Advances in Robotics: Proceedings of the 11th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines, Coimbra, Portugal, 8-10 September, 2008
[15] Nick Bostrom, Superintelligence: Paths, Dangers, Strategies, Oxford Press, 2014, pp.410
[16] Technological Singularity, Wikipedia, https://en.wikipedia.org/wiki/Technological_singularity
[17] See [15], pp.4
[18] Vernor Vinge, The Coming Technological Singularity: How to Survive in the Post-Human Era, VISION-21 Symposium, March 30-31, 1993, http://edoras.sdsu.edu/~vinge/misc/singularity.html; Ray Kurzweil, The Singularity is Near, Viking, New York, 2005
[19] Ben Goertzel, Ten Years to the Singularity: If We Really, Really Try, Humanity Press, 2014
[20] Murray Shanahan, The Technological Singularity, MIT Press, 2015

than superhuman ASI, because we will have limited time to prepare a regulatory framework while AGI is transiting into ASI. Due to this concern, in this chapter the definition of the Technological Singularity refers to the evolving nature of strong AI.

**3. The Pyramid of Robot Law**

"Micro-electrical machines" including industrial robots, airplanes, automobiles, trains, elevators and escalators have been fully incorporated into our modern society. These Action Intelligence machines (Figure 1) are well regulated under contemporary laws. In the short term a new challenge for regulators and lawmakers in order to achieve AI safety is to develop a regulatory framework which will cover the legal gap developing for emerging Autonomous Intelligence since this is a pressing issue in the timeframe of the following decade. Hence, the author introduces the term "Robot Law 1.0" to refer to a set of regulatory guidelines for intelligent machines which come with Open-Texture Risk but do not as yet have achieved consciousness or have reached the level of the Technological Singularity.

Next-Generation Robots[21] not only generate Open-Texture Risk, they also bring the new impact of affective computing to the forefront of society. Consider that in 2015, a drunken man in Kanagawa Japan entered a SoftBank as a customer. Due to a quarrel with the store clerk, he took out his anger on an intelligent machine and damaged the robot Pepper, kicking it violently. Though the clerk was not injured, the damaged robot was injured and as a result moved slower than its original interaction speed[22]. Pepper the robot has a biomorphic shape and resembles humans, and interacts socially by using its emotion, reading, and learning capabilities. The incident drew attention among legal scholars because despite how inappropriate the human-like sociable machine was treated, under the law it cannot receive any extended legal protection beyond its legal status as property, that is, as an object of law.

Has there been a governmental response to the lack of legal person status for various forms of artificial intelligence? Recently, the European Parliament proposed an independent legal personhood status for AI/robots[23]. The proposal stipulates that current policies that treat robots as property may not be sufficient to handle legal disputes involving artificially intelligent robots in the future. However, would it be prudent to consider legal personhood for artificially intelligent entities that is equivalent to rights conferred to humans? The author holds conservative views towards giving Autonomous Intelligence machines full legal personhood rights at this time. Though machines can interact with human beings via affective computing, they do not have consciousness and cannot feel real pain, anger, sorrow or happiness. Granting full legal person

---

[21] See [12]

[22] Lisa Zyga, Incident of drunk man kicking humanoid robot raises legal questions, TechXplore, https://techxplore.com/news/2015-10-incident-drunk-humanoid-robot-legal.html

[23] "DRAFT REPORT with Recommendations to the Commission on Civil Law Rules on Robotics", European Parliament, 2015, http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN

status to artificial intelligence may disturb the stable order of ethics that already exist between people and machines and the trust already established in human society. For example, suppose we granted an Autonomous Intelligence robot full legal personhood. If such a system causes harm to a person in the course of its task, it might not be liable for its wrongful, or even criminal, act.

Does it make sense to require something without self-awareness and moral standing to pay economic loss to a victim, or even incarcerate it in jail? Granting full legal personhood at the level of Robot Law 1.0 could lead to a range of legal and ethical quandaries, and inconsistencies in applying the law. Actually, in my view, what we have to protect is not the artificially intelligent robot itself, but we should require that it "project" an appropriate level of artificial empathy to human beings during daily human-robot interactions. When regulating artificial intelligence, we should avoid the negative social impacts that may happen to human beings after seeing unethical treatment given to next-generation robots. Thus, the author would like to propose granting an "extended legal protection" to Autonomous Intelligence robots with a quasi-legal personhood status called "Third Existence"[24]. If we refer to the subject/object of law as the First/Second Existence then the Third Existence manifests as neither a pure legal object, or a pure legal subject. In other words, it will be an object of law with a special legal status in order to establish a "proper" relationship between humans and robots.

What I refer to as "special" is an expedient, dynamic legal status for artificially intelligent robots. Under the premise of the Third Existence, regulators can consider cultural difference, issues of social acceptance[25], and continuing technology advancements for smart technologies, and they can refer to various concepts including corporate personhood, quasi animal personhood for Apes[26], animal welfare, non-personhood with owner liability[27], quasi personhood for unborn infants[28], and human slaves as a legal entity[29]. To treat Autonomous Intelligence robots as a Third Existence is not only to create a "shock buffer" for covering the legal gap that exists for AI personhood, but will offer benefits to policy makers by reserving core values of robot sociability with a global consensus[30].

If robots do acquire Human-Based Intelligence (HBI), one impact to the legal system will be their capability for abstract thinking, which may enable them, among others, to interpret human languages. The "formality obstacle" of Asimov's Three Laws of Robotics can be solved, but it will

---

[24] Yueh-Hsuan Weng, The Study of Safety Governance for Service Robots: On Open-Texture Risk, Ph.D. Dissertation, Peking University Law School, 2014

[25] Pericle Salvini, Cecillia Laschi, Paolo Dario, Design for acceptability: improving robots' coexistence in human society, International Journal of Social Robotics, 2(4), pp.451-460, 2010

[26] Stephen Wells, Legal Personhood for Apes, Huffpost, 2015, http://www.huffingtonpost.com/stephen-wells/legal-personhood-for-apes_b_6378486.html

[27] Richard Kelley, Enrique Schaerer, Micaela Gomez, Monica Nicolescu, Liability in Robotics: An International Perspective on Robots as Animals, Advanced Robotics, 24(13), 2010, pp.1861-1871

[28] Peter Asaro, Robots and Responsibility from a Legal Perspective, Workshop on Roboethics, IEEE ICRA, Rome, 2007

[29] See [7]

[30] Yueh-Husan Weng, Beyond Robot Ethics: On Legislative Consortium for Social Robotics, Advanced Robotics, 24(13), 2010, pp.1919-1926

lead to another concern for the violation of human-centered governance due to the consciousness (if it occurs) of artificially intelligent entities[31]. Generally, the proposal of Robot Law 1.0 will be difficult to deal with HBI self-awareness machines. The author believes that incorporating them into human society and whether or not artificially intelligent entities such as robots can receive equivalent human rights, or be recognized as a "First Existence" or subject under Law will be a hotly debated issues at the next level of regulation – Robot Law 2.0.

Except for technical challenges in machine ethics[32], some people might be curious about how the advent of robot consciousness will impact the debates about developing artificial moral agents (AMAs). Utilitarianist Peter Singer uses Jeremy Bentham's "greatest happiness principle"[33] to deduce his famous argument on equal rights consideration for animals. Suppose animals can suffer, Singer proposes that humans have an ethical obligation to avoid this undesirable outcome from happening[34]. Consciousness for an artificially intelligent entity is key here, mixed with sentience and the capability to feel pain and suffer from interactions with the real world. Hence, HBI robots should have moral standing when they cross the boundary of the Technological Singularity, and similarly I propose that this is the prerequisite in deciding whether artificially intelligent robots should deserve their own rights.

A potential concern for Robot Law 2.0 is how long it will take to go from human-level Artificial General Intelligence (AGI) to Artificial Superintelligence (ASI). Nick Bostrom believes ASI can be created soon after human-level intelligence is achieved and will result in two possible consequences – either an extremely good or an extremely bad outcome[35]. Therefore, the issue of how to develop an AI safety network in order to properly solve the "control problem" of artificially intelligent entities[36] will be another critical challenge to AI ethics. Robot Law 2.0 will not be sufficient in this regard when ASI has been created with an extremely good or bad outcome. Suppose ASI is able to enhance human life in various ways and it achieves a superior God-like status in human society. In this situation, Robot Law 2.0's proposal seeking a fairly equal relationship between human and artificially intelligent robots might seem inappropriate, even awkward. Also, there will be no need to incorporate AI safety governance into Robot Law 2.0, because many measurements relating to the control problem will not be applicable once a Superintelligence entity has been created. On the other hand, if ASI emerges as a "hurtful" entity posing dangers to humans, one possibility may be that the intelligent entity makes their own "Robot Law 3.0" and demands that human beings obey it.

The above discussion is only a brief sketch of the moral, ethical and legal issues associated

---

31 See [25]

32 Technological Challenges in Machine Ethics, Robohub, 2017, http://robohub.org/technical-challenges-in-machine-ethics/

33 Jeremy Bentham, Introduction to the Principles of Morals and Legislation. London, Payne, 1789

34 Peter Singer, Animal Liberation: The Definitive Classic of the Animal Movement, HarperCollins, 1975

35 See [15], pp.25

36 See [15], pp.155

with the three stages of artificial intelligence shown in Figure 1. If we would like to ensure the governance of AI via a regulatory framework, the author believes Robot Law 1.0 will be the most significant approach, because the time-period from Autonomous Intelligence to Human-Based Intelligence (HBI) could be several decades or even longer. However, when HBI reaches the boundary of the Technological Singularity, it may only take a short period of time before HBI evolves into Superintelligence. In other words, at the threshold of the Technological Singularity, there may not be sufficient time for humanity to develop a Robot Law 2.0.

## 4. Social System Design

Under the proposal for a Robot Law 1.0, intelligent machines should be treated as the Third Existence – objects of law with extended legal protection. Countries could then decide concrete protection measures based on their unique domestic culture and social acceptance to AI-enabled technologies. Suppose a legal object and subject occupy the very two ends of the Third Existence, so the scale could be from 0% (pure legal object) to 99.99% (very close to a legal subject, but not yet there) of the equivalence to full legal personhood based on different social values and time periods. The flexibility represented by the continuum can be the greatest common divisor of the fundamental guideline for regulators in developing suitable domestic revisions of conflicts to current existing laws and advanced intelligent technologies.

A benefit of Japan's "Tokku" special zone is that the aforementioned legal conflicts can be discovered from conducting experiments. For example, outdoor experiments using autonomous driving vehicles or humanoid robots may show the legal gap which occurs by trying to adopt advanced technologies into current road traffic regulations or other areas of human life. However, previous findings from a case study of the "robot special zone" suggests that by revising only existing laws this approach may not be sufficient to regulate advanced robotics technology controlled by artificial intelligence. The regulatory framework to AI-enabled machines should also consider other specific measures such as a "Humanoid Morality Act" and "Robot Safety Governance Act" for mitigating safety and ethical hazards[37] (Figure 1). They are role playing as "the governance in morality" and "the governance in safety" respectively. That is, we will need two specific levels of governance for AI safety and AI morality at the level of Robot Law 1.0, the "Humanoid Morality Act" and "Robot Safety Governance Act" both of which will play an important role in creating such morality/safety governance.

As AI and robotic technology continues to expand into human living spaces, the importance of the intersection between law and ethics will become more apparent and essential. Hence, we will need a "Humanoid Morality Act" to reduce the ethical gray zone and moral disputes regarding the usage of artificially intelligent robots. A special concern relating to regulating increasingly

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

smart technologies is derived from a macro perspective by looking at future human-robot interaction and their "ethical hazards", which extend from personal to societal, commercial and economic hazards, and perhaps environmental hazards as well. Examples include, but are not limited to, robot addiction, deception, and the obsolescence of jobs now performed by humans, among other things[38]. The potential demands for a "Humanoid Morality Act" can be found in the previous Pepper incident in which the robot was assaulted by a human. A Humanoid Morality Act should be at the top of a Robot Law 1.0, and will define the proper relationship between humans and robots and also direct the use of coercive power to constrain unethical applications of humanoid robotics and cyborg technologies. This approach will establish fundamental norms for regulating daily interactions between human and robot. However, it is hard to implement such regulations at this time, because it will take a period of adjustment for AI technologies to comfortably merge into our daily life. This explains why South Korea's "Robot Ethical Charter" failed at the legislative level when it was proposed in 2007[39]. In my view, the issue wasn't about the importance of the proposed topic, as regulating intelligent robots is an important topic, but the period of time they chose to enact legislation.

Even given the importance of governance in morality issues related to artificially intelligent entities, we cannot overlook the governance of safety which I propose can be done through a "Robot Safety Governance Act" – the bottom foundation of Robot Law 1.0 (Figure 1). Though AI safety has often been discussed, not all AI safety problems should be regulated by law. One example is the adaptive intelligence or machine learning capabilities of artificial intelligence, especially with regard to preventing catastrophic risks which could occur once the Technological Singularity is achieved. The story of "UK's Red Flag Laws" for steam-powered vehicles in the 19th century[40] taught us that regulators usually have a "knee-jerk" reaction towards over-regulation due to their knowledge gap which exists for technology. Based on this response, it is possible that at this time machine learning should not be regulated because it is at the cutting edge of artificial intelligence, thus it is evolving quickly. It is not likely that regulators will have equivalent domain knowledge of artificial intelligence techniques compared to professional AI programmers or robotics engineers. Therefore, the chance of over-regulation with systems using machine learning might be more pronounced than with other safety issues. Besides, the Machine Intelligence Research Institute (MIRI) has predicted the timeframe of reaching Artificial General Intelligence between 15 to 150 years[41], and the Oxford Future of Humanity Institute's survey with AI experts also revealed a diversity of professionals' opinion for the timeframe of the birth of AGI[42]. Such a "blurry" time period is not conducive for making a concrete plan for regulating AI

[38] BS 8611: 2016, Robots and robotics devices. Guide to the ethical design and application of robots and robotics systems.

[39] HB Shim, Establishing a Korean Robot Ethics Charter, Workshop on Roboethics, IEEE ICRA, Rome, 2007

[40] Red Flag Laws, Wikipedia, https://en.wikipedia.org/wiki/Red_flag_traffic_laws

[41] MIRI Website, https://intelligence.org/

[42] See [15], pp.23

safety at this point in time. In my view, this issue, safety, should be prevented by a moral consensus made by global AI communities, or as we call it "Professional Ethics".

In other words, long-term AI safety is not of significant concern for the "Robot Safety Governance Act" which refers to a regulatory framework for promising AI technologies that will be introduced into our daily lives within the following decade. The law's role in AI safety is not only about the time period in which technology comes online, but should also take priority in order to protect AI risks which present a clear and present danger to human rights. Along this line, one of the key issues is how do we govern the Open-Texture Risk from machine's autonomous behaviors which could cause physical harms to human beings?

Embodiment is another factor to consider for robot safety governance. In my view, an important issue to consider is "embodied AI"; for example, as argued by Rolf Pfeifer and Josh Bongard, embodiment is an indispensable nature of a physical entity to display intelligence[43]. At this point, the boundary of software and hardware for artificially intelligent entities are not as clear as for PCs because machines' autonomous behaviors could be generated by either its "brain's" (i.e., AI agent) decision making, its bodies (hardware) adaptive interactions with environments, or complex behaviors coordinated by both brain and body. The issue of "Body Intelligence" will bring up a host of new issues related to the safety of artificial intelligence in general, and to robot safety governance in specific. For example, how can we deal with "Modeling Error" as a potential safety gap between the machine's AI agent and hardware? Suppose its brain works normally but "Machine Fatigue" caused unwanted harmful outcomes to users, then who is responsible for the harm to humans or to property? Unless AI applications are not related to robotics, physical safety in human-robot interaction should be one of the fundamental issues of concern in AI safety.

A policy tool of importance for enacting robot safety regulations is termed Regulatory Science[44]. The US Federal Drug Administration (FDA) defines Regulatory Science as the science of developing new tools, standards, and approaches to assess the safety, efficiency, quality and performance of FDA regulated products; note that some robotic devices are considered medical devices and are regulated by the FDA. Legislators might consider making laws to restrict inappropriate interactions with robots, create personal data protection standards for humans which may be necessary given daily human-robot interaction, and regulate sales and usages of robotic technologies. It is relatively easy to develop regulations from the application side of intelligent technologies. But the design and manufacture of advanced robotics are difficult to regulate in depth because the required domain knowledge of artificial intelligence; and as stated above, technology know-how are high thresholds for non-expert law makers to overcome. Regulatory Science has been used by modern society to systematically mitigate risks from advanced technologies and these benefits create regulations in a highly technical way, so-called "Technical Norms", with examples

[43] Rolf Pfeifer and Josh Bongard, How the Body Shapes the Way We Think: A New View of Intelligence, MIT Press, 2006

[44] Regulatory Science, Wikipedia, https://en.wikipedia.org/wiki/Regulatory_science

that include current safety requirements from FDA drug regulation and UNECE motor vehicle regulation.

In Europe, regulation for robot safety is organized by two main parts as a “Directive” and a “Harmonized Standard”. The first part is a set of related EC directives which aims for harmonizing essential health and safety requirements to be applied to “machinery products.” Member states have responsibilities to incorporate machinery directives into domestic legal systems and to transform them into law to ensure the product’s free circulation in the EU market. Further, the Machinery Directive 2006/42/EC[45], the Low Voltage Directive 2006/95/EC, and the EMC Directive 2004/108/EC, are each EC directives which relate more specifically to robotics. However, these directives only define essential health and safety requirements in general applications, supplemented by a number of more specific requirements for certain categories of machinery. There is a demand to adopt extended industrial standards to improve inspection and to insure the manufacturers’ conformity to match the essential requirements. These harmonized standards provide detailed safety requirements, such as the ISO 13482: 2014 Safety Standard for Personal Care Robots[46], and are able to help verification of machine safety in a more efficient way.

In Japan, there is Consumer Product Safety Law to protect users from suffering physical harm from consumer products; the safety law, may include some types of service robots[47]. The Japanese Consumer Product Safety Law takes another safety governance approach for defining and applying several measures to product accidents. For example, a product manufacturer or importer should report to the Ministry of Economy, Trade and Industry (METI) within 10 days after he receives word of a serious accident that occurred due to use of his product. Required report items include: date, summary, name and type of the product, time and number of the manufactured or imported products, the cause of the accident, countermeasures to the accident, etc. This approach can help the government collect product accident information and then consider suitable strategies for robot safety governance.

Though there are many laws that could be used in safety regulation for artificially intelligent robots, there is a gap using them to sufficiently implement the safety governance for embodied AI, this is referred to as the “Robot Sociability Problem” discussed above. “Sociability” is the skill, tendency, or property of being sociable or socially interacting well with others. The sociability problem in artificially intelligent robotics refers to associated problems that will resemble or merge with those in other fields as robots are increasingly incorporated into human daily life. When robots become highly autonomous and are able to serve and co-exist with people, diversity values generated from robot sociability will cause the explosion of Open-Texture Risk. This problem is something current robot safety regulatory frameworks find difficult to solve. Therefore, it is inevitable that regulators address new impacts of robot technology to robot safety governance via

[45] Machinery Directive 2006/42/EC, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32006L0042
[46] ISO 13482: 2014 Safety Standard for Personal Care Robots
[47] Consumer Product Safety Law, http://www.japaneselawtranslation.go.jp/law/detail/?id=1838&re=02

embedding human-centered social values into the system and environment design process – this requires a macro level human-robot interaction or "Social System Design" approach.

Social System Design is an approach to enable integrated governance for embodied AI (Figure 2). There will be a strong demand for working ethical and legal factors into the design process of intelligent sociable robots as they are incorporated into human society. On the one hand, artificially intelligent robots should abide by moral obligations from a human-centered value system, but on the other regulators will have to consider the design of corresponding social systems in order to support their daily interactions within human living environments. Therefore, we will need an interdisciplinary approach or way of thinking about the design of artificially intelligent robots[48].

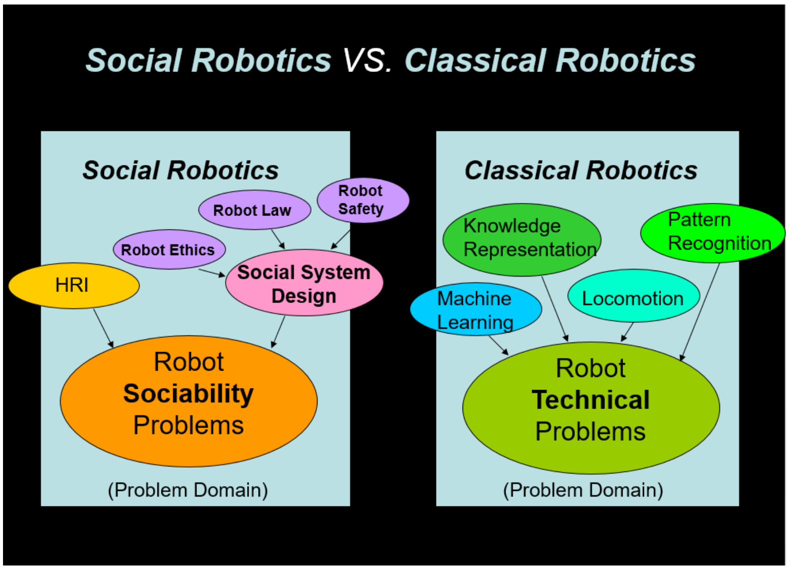


**Figure 2. Social System Design: An Integrated Approach**

The concept of a "Black Box" or Event Data Recorder is an example of an integrated governance by law and engineering. The importance of the Black Box is that its collected data can be analyzed by experts as evidence to help "Post-Safety Governance" such as liability distribution or insurance rate calculation. The major difference of a Black Box in service robot applications is that with high autonomy and sociability, robots will be incorporated into human daily life, while recording safety critical data, thus, the robot will inevitably record personal and sensitive data as well. As an example, let's say a man was killed in a house with a service robot inside. His wife (and police) wants to use the black box of the robot to know how her husband was murdered, but after she surveyed the incident, she discovered two things: one, her husband's colleague killed him, and two, a female stranger had an affair with her husband when she was not at home. The

[48] Yueh-Hsuan Weng, Towards Integrated Governance for Intelligent Robots: A Focus on Social System Design, Proceedings of the 20th International Legal Informatics Symposium – IRIS 2017, page 191-198, February 23-25, University of Salzburg

value conflict between safety and privacy as shown by this example is in conflict and something governance by law cannot solve.

"Privacy by Design" is another example of importance for regulating increasingly smart technologies. In this case, ICT manufacturers should do privacy and data protection compliance during the design and manufacturing process of their products[49]. This is an efficient way for privacy and data protection to be implemented into product design, because once personal sensitive data is made public it is difficult to make the information private again, especially once the system has been hacked. On the other hand, robot safety is similar to privacy protection, physical harms could be serious and hard to recover. This is why we will need a "Safe by Design" principle in which embodied AI systems should be created with the compliance of norms, standards or guidelines[50]. But before we touch upon the core components of the "Safe by Design" approach, we have to take a retrospective review to the development of contemporary robot safety governance.

One main focus of Aichi EXPO'05 was the future scenario of human-robot co-existence. At the EXPO, there were many service robots demonstrated to the public including Floor Cleaning Robots, Garbage Collection Robots, Security Roots, Guide Robots, Child-care Robots, and Next-generation Wheelchair Robots[51]. Importantly, the EXPO was a turning point for the Japanese government, who then decided to promote safety governance for Next-Generation Robots[52]. Afterwards, the New Energy and Industrial Technology Development Organization (NEDO) launched a project which was concerned with the practical applications of service robots and whose aim was to develop a safety governance system for physical human-robot interaction (pHRI). The project created a Robot Safety Center (RSC) in 2010 in Tsukuba science park, which has four main testing areas: (1) Traveling safety test area, (2) Collision and tip-over test area, (3) EMC test area, and (4) Mechanical test area. RSC plays the role of a certification body, the role of verification authority, and as a platform for the creation of international safety standards. Another mission of RSC is to develop evaluations for "Functional Safety" for service robots[53].

Inherent Safety ensures robot safety by removing the source of hazards, which is of low design freedom and more reliable to its users. An example of Inherent Safety is the safety fence that is used to separate humans from industrial robots. On the other hand, service robots come with high design freedom, therefore they will need another level of Functional Safety to ensure the machine safety by robots' functions. An example of a Functional Safety regulation is ISO 13482's "Virtual Fence" which sets up many different levels of safety zones, for example, robots may detect people approaching them and then decide to slow its speed and decrease its power or to stop its

[49] Privacy by Design, Wikipedia, https://en.wikipedia.org/wiki/Privacy_by_design
[50] See [2] and [38]
[51] EXPO 2005: Aichi http://www.expo2005.or.jp/en/robot/robot_project_01.html
[52] Hirochika Inoue, Robot Project for EXPO 2005, Journal of Robotics Society of Japan, Vol.24, No.2, pp. 148-150, 2006 (in Japanese)
[53] Robot Safety Center, http://robotsafety.jp

task altogether to realize safety via these functions.

As we enter the era of artificially intelligent robots, we might need to consider safety hazards alongside new hazards in order to ensure the safety of physical human-robot interaction. Autonomous Intelligence machines and Open-Texture Risk are both mixed with safety and ethical hazards, therefore, only removing the source of hazards in advance or by ensuring the safety with functions are not enough. We will need a mechanism called "Safety Intelligence" to help avoid unwanted safety behaviors by artificially intelligent machines in real time[54]. At least in theory, Isaac Asimov's Three Laws of Robotics is an ideal paradigm of Safety Intelligence. It is inevitable that legal schemes consider authorizing machines with a safety-oriented autonomy in unstructured environments when they co-exist with humans. A key challenge for legislators and judges will be to design a safety governance protocol that can properly coordinate inherent safety, functional safety, and safety intelligence into one regulatory scheme. These protocols will need to be put into the basket of safety intelligence should emerging AI technologies which have concrete applications be introduced into the market. They include, but are not limited to, machine learning, computer vision, simultaneous localization and mapping (SLAM), nature language processing (NLP), and haptics sensing.

Additionally, we might be able to define legal requirement for computer vision, such as letting computer vision recognize (by nature of the algorithms used) a set of non-verbal cues or body language or SLAM that teaches an artificial intelligence run machine to keep a proper social distance from people, or to avoid the negative side effects from a cleaning robot with RL agents[55]. On this point, regulators only have to consider how to properly control the Open-Texture Risk from emerging AI technologies. Hence, they can be released from a series of philosophical debates on smart robots and their artificial morality.

The specifics of the legal framework of Robot Law 1.0 is presented on the upper and lower sides of the "governance in morality" and "governance in safety", and its universality is between the two in the form of "revisions" (See Figure 1). Over the past few years there have been many examples of conflicts between existing laws and robotics, such as: bipedal humanoid robots or unmanned vehicles immediately faced with restrictions based on road traffic laws when they leave the laboratory to enter the real world. Further, other issues of interest to artificial intelligence which will lead to privacy protection and other issues will be the difficulty in judging tort liability resulting from an accident caused by autonomous robots; the use of sensor equipped smart unmanned aerial vehicles which will lead to privacy protection issues; the issue of lethal military robots which will cause international humanitarian law disputes, and so on. Finally, the law must consider appropriate amendments of current law and statutes to promote the integration of robotic

---

[54] Yueh-Hsuan Weng, Chien-Hsun Chen, Chuen-Tsai Sun, Toward Human-Robot Co-Existence Society: On Safety Intelligence for Next-Generation Robots, International Journal of Social Robotics, 1(4), 2009

[55] Dario Amodei, Chris Olah, Steinhardt, Paul Christiano, John Schulman, Dan Mane, Concrete Problems in AI Safety, arXiv preprint arXiv:1606.06565, 2016.

science and technology into society. In the embryonic period of human-robot co-existence in society, "revisions" involve only a small part of many statutes, such as road traffic law, tort liability law, and International humanitarian law. But with robot technology becoming smarter, given its scope of expansion in society and generally the capabilities of robot technology itself, the development and progress of the "revised laws and regulations" is bound to extend the scope of the regulation of smart technologies to other laws, such as intellectual property law, criminal law, and even the Constitutional law of most jurisdictions.

**5. Conclusion**

The importance of regulating artificial intelligence can be shown by increasingly smart robotics which takes the form of an embodied medium for AI agents thus allowing the agent to physically interact with human beings. In this chapter, the author described the associativity between AI and robotics, and discussed the evolution of a legal framework for dealing with AI in different stages of its development. Considering the strong connectivity between AGI and ASI, I propose that we shall seriously consider using Robot Law 1.0 to improve safety governance for embodied AI. Further, Social System Design is not only a measure for establishing AI's social status in human society, but also a moderator to keep the balance between management, technology, professional ethics, and law among AI safety governance. This integrated governance approach will ensure the future regulation of artificially intelligent entities in a safe and effective manner and in particular allow for human-robot co-existence.