

Yueh-Hsuan Weng

Towards Integrated Governance for Intelligent Robots: A Focus on Social System Design

As intelligent robots become increasingly common in human society, it will be essential to begin incorporating ethical and legal factors into their design process. Hence, legislators and policy makers should learn from the interdisciplinary concept of «Social System Design» in order to better regulate social robots and their attendant AI-driven risks. This problem has many facets, and as such suggests an interdisciplinary approach towards thinking about the design of emerging intelligent machines.

Category: Articles

Region: Japan

Field of law: AI & Law; Robotic

Citation: Yueh-Hsuan Weng, Towards Integrated Governance for Intelligent Robots: A Focus on Social System Design, in: Jusletter IT 23. November 2017

Contents

1. Introduction
2. Regulation of Unknowns: A Tendency Towards Over-regulation
3. Towards the Integrated Governance of Intelligent Robots
 - 3.1. «Tokku» RT Special Zone: A Tool for Deregulation
 - 3.2. Certification
 - 3.3. Professional Ethics
 - 3.4. Designing Robot Sociability
4. Conclusion

1. Introduction

[Rz 1] Since the advent of the Industrial Revolution, steam-powered and micro-electrical machines have become more and more common, expanding their role in society up until today. At this point, human beings have already co-existed with these machines for more than two centuries, and contemporary law has developed preventionary regulatory frameworks to ensure the safety of machines, including automobiles, railway systems, elevators and industrial robots, amongst others. However, we are witnessing the beginning of a new era of intelligent robots. These next generation robots will be capable of adapting to complex, unstructured environments and interacting with humans to assist them with tasks in their daily lives. Unlike heavily regulated industrial robots that toil in isolated settings, next generation robots will have relative autonomy, which raises a number of new ethical, legal, and safety issues.

[Rz 2] A decade ago, Japanese Ministry of Economy, Trade and Industry (METI) began developing a series of business, innovation and safety policies for robotics. Its latest comprehensive policy guideline for regulating robotics, known as «New Robot Strategy», is a five year grand strategic plan that aims to promote Japan's competitiveness through both regulation and deregulation. The guideline encourages the development in the areas of «Artificial Intelligence (both data-driven and brain-like),» «Sensing and recognition technology,» «Mechanism, actuator, and control technology,» and «OS and middleware.» These core technologies will be crucial for developing «Next Generation Robots» and could be applied in a wide range of real world sectors such as manufacturing, service, nursing and medical, construction and agriculture.

[Rz 3] With regard to the subject of law and robotics, the Robot Revolution Realization Council had systematically inspected potential problems within Japan's legal system, leading to revision of many existing regulations, including «Radio Law,» «Pharmaceuticals and Medical Devices Law,» «Industrial Safety and Health Act,» «Road Traffic Law,» «Road Transport Vehicle Act,» «Civil Aeronautics Act,» «Control Law of Unjust Access,» «Consumer Products Safety Act,» «ISO 13482 Safety Standard for Life-supporting Robots,» and the «Industrial Standards Law.»

[Rz 4] In accordance with these realizations, the Robot Revolution Realization Council proposed the «Implementation of Robot Regulatory Reform» as a guideline. This document outlines two strategies for regulatory reform. The first eases current regulations by creating new legal frameworks or utilizing the environment. The other strategy is to establish a new legal framework from the consumer protection perspective. In addition, field testing for robots is an essential part

of deregulation, because it can help regulators and manufacturers to discover many unexpected risks during the final stage prior to a machine's real world deployment.¹

[Rz 5] In the United States, the Obama White House published a report on «Artificial Intelligence, Automation, and the Economy» in December 2016 which suggests three main strategies for improving the US economy with regard to the impact of AI-driven automation.² Across the Atlantic, European Parliament Rapporteur Mady Delvaux published a draft report with recommendations for the Commission on Civil Law Rules on Robotics in May 2016 which urges European Commission to create a robust EU-wide framework on standard levels of safety and security for emerging AI and robotics technologies, including a voluntary Code of Ethical Conduct, a new EU agency for AI and robotics, and a specific legal status for electronic persons, amongst other things. Delvaux's report was approved by the European Parliament's Legal Affairs Committee in January 2017 and is expected to be voted on at the full meeting of the European Parliament in February 2017.³

2. Regulation of Unknowns: A Tendency Towards Over-regulation

[Rz 6] Pepper the robot, developed by SoftBank and manufactured by Foxconn, is able to interact socially with human beings using on its emotion reading and learning capabilities facilitated by cloud computing. Pepper is a highly intelligent machine that can read human emotions, respond to inquiries, and interact with human beings. It even has a biomorphic shape resembling that of humans. However, recently a Pepper robot was attacked by a drunkard in Japan. Was there anything wrong with Pepper's emotion reading function, or was the man at fault?

[Rz 7] According to the Japan Times, on a Sunday morning in Yokosuka, Kanagawa, a drunken man entered a local SoftBank store and kicked a Pepper robot stationed there. The man was soon arrested by the police. He has admitted to damaging the robot, claiming that he did not like the attitude of a store clerk. Though the clerk was not injured, the damaged robot now moves slower than its original interaction speed.⁴

[Rz 8] This incident is worth discussing, as there may be clues for us to think about emerging issue around robots coexistence with society at large. The incident has received immense scrutiny from the public, raising the issue as it does of whether the human-like sociable machine was inappropriately treated. If the object involved had been an ATM or a vehicle, the moral impact would certainly have been much smaller. But an evolved set of ethical principles for sophisticated and intelligent machinery like Pepper has yet to be developed.

¹ THE CABINET OF JAPAN, Draft outline of the robot revolution realization council. Japanese Prime Minister and his Cabinet, 2014 (<http://www.kantei.go.jp/jp/singi/robot/> [all Internet sources accessed on 7 February 2017]).

² The White House, Artificial Intelligence, Automation, and the Economy. US Executive Office of the President, 2016 (<https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>).

³ EUROPEAN PARLIAMENT, Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), 31 May 2016 (<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN>).

⁴ Drunken Kanagawa man arrested after kicking SoftBank robot, The Japan Times 7 September 2015 (<http://www.japantimes.co.jp/news/2015/09/07/national/crime-legal/drunken-kanagawa-man-60-arrested-after-kicking-softbank-robot-in-fit-of-rage/#.WJV1Xv195Pb>).

[Rz 9] A lesson can be learned from 19th century, the end of the era of horse-drawn transportation. The origin of human and horse co-existence can be traced to horses ridden by nomadic herders in Central Asia 5'000 years ago. With the accompanying inventions of bits, collar harnesses and coaches, horse-drawn transportation gradually became a dominant method of land transportation.⁵ ERIC MORRIS also pointed out that horses were absolutely essential for the functioning of 19th century cities in Western countries, mainly for personal transportation, freight haulage, and mechanical power.⁶ Around this time, however, the rise of steam engine technology brought new possibilities for personal transportation. Richard Trevithick invented the world's first self-propelling passenger-carrying vehicle, called the «London Steam Carriage», in 1803, and later the «Stockton and Darlington Railway», the world's first public railway to use steam locomotives, which opened in 1825. Unlike steam locomotives, which have their own independent railway networks, steam powered automobiles had to be operated and tested in human living areas, in particular on public roads. It raised many new social concerns, such as how to limit the speed of self-propelled vehicles, and how to ensure pedestrians» safety. For example, if a horse carriage meets a steam car face to face, what should happen next? And how could one prevent a horse from being startled by a steam car's emitted vapor?

[Rz 10] In the mid-19th century, the UK Parliament introduced legislation called the «Red Flag Laws» to regulate steam-powered automobiles. These laws were implemented between 1861 and 1896. However, the regulations were relatively cautious and conservative. For example, the law asked for at least three people to be in charge of the automobile's operation: the driver, an engineer, and a flagman. It was proposed that the flagman should walk slowly preceding the moving car by no less than 60 yards, and wave a red flag or carry a lantern to warn horse riders or pedestrians to ensure safety.

[Rz 11] Unfortunately, the effects of this regulation were disappointing. Aside from holding a red flag, other strict laws including a speed limit of 2–4 mph (3–6 kph) and additional toll fees for steam-powered vehicles using non-cylindrical wheels were implemented, though some were eventually overruled. Though regulation can effectively reduce risks from emerging new technologies, it can also prevent further innovation as well as retard industrial growth. Researchers believe the adoption of the Red Flag Laws can explain why UK's automobile industry fell behind that of Germany and France. Surprisingly, however, the world's first steam powered passenger-carrying vehicle did come from the UK.⁷

[Rz 12] With regard to the Pepper incident, a humanoid robot is recognized as an «Object of Law» under the current Japanese legal system. Therefore, it is not possible to apply Article 204 (Injury) of Japanese Penal Code. However, the man could be charged under Article 234-2 (Obstruction of Business by Damaging a Computer) or Article 261 (Damage to Property). As for civil law, based on Article 709 (Damages in Torts), Pepper's owner, SoftBank, can claim economic compensation to the man for any damages resulting from the attack on the Pepper robot.

⁵ The Educational Programming Guide for Going Places, ExhibitsUSA, a national division of Mid-America Arts Alliance, 2007 (<http://parkcityhistory.org/wp-content/uploads/2012/04/Teacher-Background-Information.pdf>).

⁶ ERIC MORRIS, From horse power to horsepower, Access Magazine 2007, No. 30.

⁷ Locomotive Acts, Wikipedia (https://en.wikipedia.org/wiki/Locomotive_Acts; <https://ja.wikipedia.org/wiki/%E8%B5%A4%E6%97%97%E6%B3%95>).

[Rz 13] Following the analysis above, it is apparent that we should consider addressing the impact of new regulations on service robots. Under the current legal system, service robots are merely property or a «second existence». This is not enough to address safety and moral risks in regards to human-robot co-existence. In other words, a new perspective for regulation should be established under the premise that service robots ought to be recognized as a «the third existence» legal entity, where robots are still an object under the law, but should have a special legal status different from that of normal machines.⁸ The difficulty of implementing new regulation for service robots, however, is similar to the case of regulating steam powered cars in the 19th century. It's a «Regulation of the Unknown». On the one hand, such machines could have lethal consequences for human beings without proper regulation. On the other hand, it can be difficult for regulators to keep up with the progress of advanced technologies. If there is a tendency towards over-regulation, similar to the case of the steam powered cars, there will be problems.

3. Towards the Integrated Governance of Intelligent Robots

[Rz 14] ROLF PFEIFER has argued for the indispensable nature of a physical entity to intelligence in his thoughts on «embodiment».⁹ At this point, robotics takes the form of an augmented reality medium for AI agents to be able to physically interact with human beings. If we say that Cyber Law is a set of rules for regulating human beings» interactions with the virtual world, then we might refer to Robot Law as a set of specific norms which aims to mitigating risks inherent in AI agents» behaviors and their consequences in the real world. Hence, we will need to develop a regulatory framework not only for AI, but including its accompanied field of advanced robotics, or «AR».

[Rz 15] In order to avoid the establishment of a new set of «Red Flag Laws» in the era of intelligent machines, we have to seek an objective way to help reduce the tendency for over-regulation of AI technologies. Modern society had gradually developed a risk governance system based on «regulatory science» to systematically mitigate risks from advanced technologies such as is seen in new drug development. According to the FDA, the definition of regulatory science is the science of developing new tools, standards, and approaches to assess the safety, efficiency, quality and performance of FDA regulated products.

[Rz 16] That said, our findings from an empirical case study of the Tokku special zones suggests that mere revision of existing laws to regulate advanced robotics technology is not enough. We will need other specific measures such as a «humanoid morality act» and «robot safety governance act» for mitigating these safety and ethical hazards. Relevant examples include current FDA drug regulation and UNECE motor vehicle regulation. These regulations are used to present, in a highly technical way, so-called «Technical Norms». In other words, the governance systems for embodied intelligence will need an integrated approach to regulating AI risks at the front end of the robot design process.

⁸ YUEH-HSUAN WENG/CHIEN-HSUN CHEN/CHUEN-TSAI SUN, Toward the human-robot coexistence society: on safety intelligence for next generation robots, *International Journal of Social Robots* 2009, Volume 1, Issue 4, pp. 267–282.

⁹ ROLF PFEIFER/JOSH BONGARD, *How the Body Shapes the Way We Think: A New View of Intelligence*, MIT Press, New York 2006.

[Rz 17] Based on this unitary viewpoint to emerging robotics» AI governance we should focus on intelligent machines which are accompanied by Open-Texture Risk but are yet to be Self-aware, and might be called a «Third Existence», as mentioned above. My proposed Social System Design for embodied Intelligence can be divided into four main parts: (1) Deregulation, (2) Certification, (3) Professional Ethics, and (4) Designing Robot Sociability.

3.1. «Tokku» RT Special Zone: A Tool for Deregulation

[Rz 18] Looking at integrated governance of AI ethics, we can first consider «Deregulation» in the context of the «Tokku» RT (Robotics Technology) special zone. A special area such as this can help regulators and manufacturers discover unexpected risks during the final stage prior to a machine's real world deployment. Originating in Japan, the RT special zone has only been around since the beginning of 21st century but in that time there have already been many other special zones established in places like Fukuoka, Osaka, Gifu, Kanagawa and Tsukuba. As the development of robotics advances and its prevalence in society grows, the importance of special zones of interface for robots and society will be more apparent.¹⁰

[Rz 19] In the short term, a special zone is a kind of special measure to enhance the competitiveness of domestic RT industry, but in the long term it can also function as a «Shock Buffer». An RT special zone contributes to: (1) ensuring machine safety, (2) preventing high litigation risks, and (3) easing radical ethical disputes. These could all be helpful for regulators developing «Robot Law» in other countries. While this legal institution in Japan does not refer to the debate on issues like the rights of robots or whether robots ought to be recognized as a subject under the law, it does discuss public regulation of the design, manufacture, sale and usage of advanced robotics. A possibility could be developing a «Robot Safety Governance Act», which would be an extension of current machine safety regulations. These technical norms derived from the foundation given by the «Robot Law» will ensure the safety of a new human-robot co-existence.

[Rz 20] As robotic technology continues to expand into human living spaces, the importance of these legal and ethical concepts will become more apparent and essential. A «Humanoid Morality Act» can reduce the ethical gray zone and moral disputes regarding the usage of service robots. This «Humanoid Morality Act», which should be a fundamental part of «Robot Law», will define the proper relationship between human and robots and direct the use of coercive power to constrain unethical applications of humanoid robotics and cyborg technologies. This will establish fundamental norms for regulating daily interactions between human and robots. Clues regarding the potential demands for a «Humanoid Morality Act» can be found in the Pepper incident.

[Rz 21] In addition, robot ethics and legal regulation should not always be in parallel, because from the regulation perspective, robot law is a union of robot ethics and robotics. Through the deregulation system, we will be able to discover and evaluate potential AI risks inside various forms of human and robot daily interaction.

¹⁰ 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 Robots* 2015, Volume 7, Issue 5, pp. 841–857.

3.2. Certification

[Rz 22] Physical safety in human-robot interaction is not merely the fundamental issue of concern in robot ethics, but also an important factor in product liability. Therefore, in many countries robot manufacturers will have to send their products for safety certification in order to match expectations from the market and also to help clarify potential liabilities to relevant stakeholders. In 2009, Japan's New Energy and Industrial Technology Development Organization (NEDO) launched a project concerned with the practical applications of service robots which aims to develop a governance system for physical human-robot interaction (pHRI). The project created a robot safety testing center inside Tsukuba RT special zone to provide safety certification for personal care robots, which is defined under ISO 13482.¹¹

[Rz 23] However, there is another type of AI hazard in daily human-robot interaction known as «ethical hazards». ALAN WINFIELD, who is a member of BS 8611 – the world's first standard highlighting the ethical hazards of robots, notes that BS 8611 takes a macro perspective in looking at future human-robot interaction. It covers many examples of ethical hazards, which go all the way 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.

[Rz 24] CHRISTOPHER HARPER and GURVINDER VIRK argued that when dependability is a legal requirement, robots will require certification before they can be put into service. Standards are crucial in the process of certification because they capture the consensus of the best practice of safe system behavior and design methodology. As we enter the era of intelligent robots, we might need to consider safety hazards alongside new ethical hazards in order to ensure the physical safety factors inherent in the use of intelligent robots. If so, then the current legal requirement of dependability might not be enough to fulfill a certification.¹²

3.3. Professional Ethics

[Rz 25] Artificial intelligence and robotics are highly specialized scientific disciplines, and as such it is not likely that regulators will have equivalent domain knowledge to professional AI programmers or robotics engineers. Hence, a crucial governance tool for AI safety is the promotion of professional ethics inside the global AI and robotics communities. In 2006, the European Robotics Research Network (EURON) had published its «Roboethics Roadmap», a collection of articles outlining potential research pathways, and speculating on how each one might develop. Around a decade later, IEEE Standards Association's global initiative for ethical considerations in artificial intelligence and autonomous systems published their guideline, called «Ethically Aligned Design» in December 2016. The main purpose of IEEE global initiative is to develop a new framework of ethical governance for artificial intelligence system design complete with future norms and standards. Compared to EURON Roboethics Roadmap, one major difference is that IEEE Ethically Aligned Design not only encourages professional ethics in AI, but also con-

¹¹ A Guidebook on the Project for Practical Applications of Service Robots, New Energy and Industrial Technology Development Organization (NEDO), Tokyo 2011.

¹² CHRISTOPHER HARPER/GURVINDER VIRK, Towards the Development of International Safety Standards for Human Robot Interaction. International Journal of Social Robots 2010, Volume 2, Issue 3, pp. 229–234.

sider embedding norms and values in artificial intelligence systems which jumps into the realm of machine ethics.¹³

3.4. Designing Robot Sociability

[Rz 26] An approach to address the expanding legal gap caused by the AI revolution is to overcome the «Robot Sociability Problem», which refers to associated problems that will resemble or merge with those in other fields as robots are increasingly integrated into human society. Sociability is the skill, tendency or property of being sociable or social, or of interacting well with others. This ability is very important to human beings; compared with human sociability, robot sociability is clearly artificial, and therefore the property and relationship between humans and robots can be totally controlled by humans. Designing Robot Sociability can be divided into two aspects. The first is a micro perspective or Human-Robot Interaction, which covers various unspoken rules in human-robot daily interactions. For example, for a service robot in human living environments it is important for it to know proximities or how to keep a proper distance from different people or situations. The second is the macro perspective, where the aim is to decide what kind of ethics, policy and law can be applied into autonomous robots, called Social System Design.

[Rz 27] 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, intelligent robots should abide by moral obligations from a human-centered value system, but on the other hand, regulators 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 way of thinking about the design of intelligent robots.

[Rz 28] The first step of Social System Design could be to create a regulatory taxonomy of AI, which does not only benefit regulators to define regulated objects and to plan possible strategies for AI governance, but also helps to deal with issues such as accident responsibility distribution. Here the example is Robot Law 1.0.

[Rz 29] Neurologists view the human brain as having 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».¹⁴

[Rz 30] From an AI viewpoint, the biomorphic equivalents of the three layers are action intelligence, autonomous intelligence, and Human-Based Intelligence. Action intelligence functions are analogous to nervous system responses that coordinate sensory and behavioral information, thereby giving a robot the ability to control head and eye movement, move spatially, operate machine arms to manipulate objects, and visually inspect its immediate environment. Autonomous intelligence refers to capabilities for solving problems involving pattern recognition, automated scheduling, and planning based on prior experience. Such behaviors are logical and programmable, but not conscious. They have remarkable abilities to perform specific tasks according to their built-in autonomous intelligence. However, they cannot make decisions concerning self-beneficial actions or decide what is right or wrong based on a sense of their own

¹³ ALAN WINFIELD, *Ethically Aligned Design*, Robohub 2016 (<http://robohub.org/ethically-aligned-design/>).

¹⁴ PAUL D. MACLEAN, *The Triune Brain in Evolution*, Plenum, New York 1990.

value. At the third level is Human-Based Intelligence – higher cognitive abilities that allow for new ways of looking at one’s environment and for abstract thought, also referred to as «mind» and «real intelligence». Since a universally accepted definition of human intelligence has yet to emerge, there is little agreement on a definition for Human-Based Intelligence. The upper level intelligence is Superintelligence. Nick Bostrom defines superintelligence as «an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills».¹⁵

[Rz 31] Action intelligence machines are well regulated under contemporary laws. Hence, the author would like to skip this part. A more pressing issue for the following decade will be Next-Generation Robots or autonomous intelligence machines which come with Open-Texture Risk but do not yet have self-conscious or pass the line of Singularity. In other words, it is a Third Existence. Neither is it a pure legal object, nor a pure legal subject. From the previous case in Japan, we understand that there exists a gap in existing law. Therefore, in the era of intelligent robots, we will need a new legal framework Robot Law 1.0 for Third Existence in order to keep a proper relationship between humans and robots. Some people might be curious how the advent of robot consciousness will impact the legal system. If robots do acquire human-level intelligence, I believe that whether or not robots can receive equivalent human rights, or be recognized as a «First Existence» or Subject under Law will be hotly debated issues in the next level, Robot Law 2.0. According to experts from Oxford FHI, someday in the near future, Robot AI could surpass human-based intelligence and reach the next level – Superintelligence. I predict that when AI reach this level there will be a structural change to a corresponding Robot Law 3.0. One possibility may be that the intelligent entity makes their own Robot Law and asks that human beings obey it.

[Rz 32] This is only a brief sketch of future Social System Design for embodied intelligence. If we would like to ensure AI Safety via a regulatory framework, I believe Robot Law 1.0 will be the most significant, because the time period from autonomous intelligence to human-based intelligence could be several decades or even longer. However, when human-based intelligence reaches toward the boundary of Singularity, it may only take a quite short period of time before human-based intelligence will evolve into Superintelligence. In other words, there may not be sufficient time to develop a Robot Law 2.0 to ensure AI Safety.

4. Conclusion

[Rz 33] One of major concerns associated with AI risk regards physical interaction with human beings. Therefore, embodiment will be a key factor in AI Safety. In this case, we might pursue an integrated approach of AI governance considering deregulation, certification, professional ethics, and designing robot sociability. Finally, with regard to social system design for embodied intelligences, some major challenges are likely to be faced in developing Robot Law 1.0: (1) Social Robotics is important for the implementation of a human-robot co-existence society. However, how exactly we ought to keep a balance between Human-Robot Interaction and Social System Design is yet to be determined; (2) Looking at the regulation of as yet unknown new technologies, we may need to consider developing a corresponding «deregulation» system, with the Tokku Special

¹⁵ NICK BOSTROM, *Superintelligence*, Oxford University Press, Oxford 2014.

Zone as an example; (3) How to realize the «Ethics by Design» principle in the design, manufacture, and use of autonomous systems remains to be seen; (4) In the near future a priority concern will be Third Existence machines. Should we aim to prepare a short term AI Policy governing its co-existence with human beings?

YUEH-HSUAN WENG, Assistant Professor, Frontier Research Institute for Interdisciplinary Studies (FRIS), Tohoku University (starting march 2017); Aramaki aza Aoba 6-3, Aoba-ku, Sendai 980-8578, JP; weng.yuehhsuan@gmail.com; <http://www.robolaw.asia>.