MAKING OF SOUTH KOREAN ROBOT ETHICS CHARTER: REVISED PROPOSITION IN 2018

Young Lim Choi[1], Eun Chang Choi,[2] Dang Van Chien[3], Tran Trung Tin[4] and Jong-Wook Kim[5]

Department of Electronic Engineering, Dong-A University

 # RS906 37, Nakdong-daero 550beon-gil,
Saha-gu, Busan, Republic of Korea

Artificial Intelligence and Robotics Lab
http://deas.donga.ac.kr

Abstract   The aim of this paper is to give a gist of the 2018 Robot Ethics Charter in South Korea that provides design principles for controlling intelligent robots and maximizing human safety. The Charter, shaped by the Korea Robotics Society, is meant to be applicable at the whole stages including design, manufacturing, management and usage of artificial intelligence and robotics in a way that ethical requirements are being recommended to each actors. The newly available Charter, released in 2018, contains its values and core principles.

1. Introduction

In his seminal paper published in 1950, Alan Turing proposed a fairly vague question "Can machine think?" [1] At the time, he demonstrated that any recursively computable function can be computed in finite time by a simple sort of symbol-manipulating machine. Almost seven decades later, AI programmers garnered more reasons for saying "machines can think". Robotics or intelligent machines are being imbued with the effective capacity to learn. The development of high-performance hardware such as GPU and highly sophisticated computer algorithms has allowed AI system and robotics to learn unsupervised data sets that leads to autonomous decision-making of a series of desirable moves to complete a given task. So, we are moving towards an era in which intelligence robotics recognizes human language and voice inflection and responds appropriately.

As a part of AI-enabled autonomous systems, robotics are specifically designed to achieve a particular task. Basically, robotics are expected to perform repetitive or extremely dangerous tasks, and thereby improving the quality of human life. Nonetheless, the functions of robotics are not always guaranteed to produce acceptable results without physical harm or unwanted damage. For instance, although autonomous systems have learned from examples that may cause dangerous errors in the real world application. As robotics are being deployed widely, there are increasing concerns about the dark side of robotics i.e., unexpected misbehavior, hazardous actions, or mal-function that are linked to safety for humans.

Accordingly, robotics developers are certainly convinced that society's acceptance of intelligence machines depends on whether intelligent machines can be programmed to make decisions in ways that ensure human safety, and fit in with social norms. [2] So, technology ethicists, who seeks to understand the moral implications of AI development, as well as robotics developers attach great importance to the design principles for safety requirements in

---

[1] Artificial Intelligence and Robotics Lab, Dong-A University

[2] Affiliate Fellow, The Information Society Project, Yale Law School

[3] Artificial Intelligence and Robotics Lab, Dong-A University

[4] Artificial Intelligence and Robotics Lab, Dong-A University

[5] Artificial Intelligence and Robotics Lab, Dong-A University

 * Corresponding Author < kjwook@dau.ac.kr>

intelligent systems. In this sense, the Korea Robot Society suggested the Robot Ethics Charter in September 2018 in order to curb undesirable outcomes with robotics. The Charter would work as a guideline in tandem with the design principles if it should be implemented.

2. Background

Back in 2007, the Robot Ethics Charter Establishment Committee of Korea, an advisory body under Ministry of Commerce Industry and Energy (MCIE), has sketched out a preliminary draft of Robot Ethics Charter.[3] It was an ethical code to prevent robots abusing humans, based on Isaac Asimov's ideas on 'Three Laws of Robotics'. The draft, however, was not proceeded to deliberate at the Industrial Development Council since then. It couldn't reach a consensus because it was basically regarded as a premature proposal. More than a decade later, a newly-formed Charter was shaped by Robot Ethics Research Committee, chaired by Professor Jong-Wook Kim, of the Korea Robotics Society (KROS). The Charter was presented at the Robot Ethics Forum in September 2018 in Seoul. Unlike its technological setting of 2007, robotics industry is developing more complex intelligent robotics. Intelligent robots are equipped with capabilities in sensing, interaction, problem solving, and learning, stochastic models in robotics make their actions are unpredictable. Intelligent robots learn to detect the right patterns and act exactly according to their input. But these systems can be fooled in ways that humans wouldn't be. 'Artificial stupidity' or deliberate errors cannot be perfectly screened through the training phase. Thus, the ethical dilemmas of robotics is getting serious.

Meanwhile, KAIST, a leading university in South Korea, faced backlash over opening AI lab on account of 'killer robot' fears. It transpired just a few months before the release of the new proposition. Global research community in AI and robotics signed an open letter opposing its participation in an autonomous-weapons race.[4] The letter initiated an alarming outburst among researchers worldwide. It, however, was not the case in that the AI lab just launched with focus on defense technologies. There were literally no lethal, autonomous robots that can carry out targeted strikes on humans. Conversely, there was only research scheme with a defense contractor to improve existing defense systems. For instance, AI algorithm for an unmanned submarine's navigation, one of research agendum of the lab, could minimize the possible human casualties. In order to provide assurance, the president of KAIST had to expound on "no intention" to engage in development of lethal autonomous weapons systems. The statement stressed the principle of 'meaningful human control' over autonomous systems which has been imprinted on the public's mind. The aftermath of the boycott letter significantly highlighted the necessity of ethical standards in robotics development at any rate.

3. Purpose and Preparation

The proposed Charter of 2018 contains general principles and design standards, so that robots will serve the whole human beings' happiness in the future society. The purpose of the Charter is to present a guidelines to be considered in the phases of design, manufacture, supply, usage and management of AI and robotics. Therefore, a wide range of industrial manufacturers, public organizations, service operators, and companies related to robotics are encouraged to comply with the Charter's ethical code. But its nature is not considered legally binding.

The Charter was intended to offer a guideline with a relatively plain account, so that it might develop the broad social consensus on the usage of robots. During the preparation process, the Charter collected opinions from ethicists, legal scholars, robotics industry and software engineers. It also referred to the global trend with regards to ethical standards in robotics and artificial intelligence. The Charter contains a realistic standard that can be applied to all actors with a nonbinding ethical code. Furthermore, the proposed Charter is schedule to take more feedbacks from various strata of society into consideration. On the flip side, as of June 2019, the proposed Charter has yet to be officially adopted as an 'Intelligent Robot Charter' by the Korean Robot Industry Policy Deliberation Committee housed in the Ministry of Trade, Industry and Energy (MOTIE) of South Korea.[5]

3. The Fundamental Values

In the Charter, 'Robot' is defined as machine with mechanism that autonomously operates by recognizing the external environment and judging the situation. 'Artificial Intelligence' is a technology that implements the functions of human intelligence such as cognition, learning, reasoning, judgment, memory, and natural language processing by software design. The Charter consists of three fundamental values, and five principle of practices.

3.1. Protection of Human Dignity

The aim manufacturing of AI·Robotics and their actions cannot be used to incur harm or instrumentalize humans. AI· Robotics should be designed, manufactured, supplied, used, and managed to protect fundamental human rights— individual freedom, privacy, personal data, and safety. These technologies and services are prohibited to treat humans merely as means, and should be developed and used in ways that respect and protect human dignity. The AI·Robotics should not discriminate against people based on gender, age, disability, race, religion or nationality in the phases of design, manufacture, supply, usage, management of it.

3.2. Pursuit of Public Good

AI·Robotics should be designed, manufactured, supplied, used, and managed in a way that improve human public welfare. AI based services should be designed, manufactured, supplied, used, and managed for the benefit of the greatest number of people, and should enhance the capacity of them. And these technologies and services should strive for equality ensuring the accessibility of the marginalized and vulnerable groups. AI·Robotics should be used within the range of pursuing the private interests unless the public interests is significantly greater than the private interests.

3.3. Pursuit of Happiness

AI·Robotics are expected to serve to improve the quality of human life and promote happiness of humans. AI·Robotics are not meant to be an autonomous status in relation to humans, but only treated an instrument. These technologies and services are nothing more than a means to benefit humans. Although AI·Robotics can be designed to seem like having intelligence or emotion, the reality of the inner mechanism must be clearly notified to the users.

4. The Applicable Phases and Actors

Provided that the Charter is a proposed guideline for the actors involving in the various phases of the cycle, it is not accompanied with any obligation or regulatory enforcement. It is notable that the Charter is based on the premise of the overall applicability. The principle of practices can be implemented in designing, manufacturing, service providing, using, and managing of robotics. Thereby the Charter imposes the package of ethical requirements to be considered at each phase of the cycle. The actors of Charter are in all stages – design, production, supply, use, and management of robotics. The actors are divided into three parts: 'robotics manufacturer', 'service provider', and 'user'. The actors should act ethically throughout the stages of design, production, supply, use, and management of robotics.

4.1. Manufacturer

This category encompasses any person, organization, or company that develops, manufactures, sells, manages, researches, designs, produces or provides any kind of element technology, integration technologies, services and products, or robotics theories related to its design or controlling in the field of hardware or software for robotics including computer algorithm, robotics parts, components, controllers, or Human-Robot Interaction (HRI) etc. For example,

in case of elder care robots, it would include gerontologist, psychologists, ergonomists, philosophers, etc.

4.2. Service Provider
A person, agency, company or organization that creates added value to the ready-made AI·Robotics by selling, supplying, installing, managing by system integration, providing or operating various services according to market demand. For example, a management company that provides a drone or guard robot service; a building cleaning service company that uses cleaning robots, etc.

4.3. User
Any person, organization or company that directly or indirectly uses the services that artificial intelligence and robots can provide. For example, citizens who purchase or use drones; a group of people exposed to drone flights; teachers and students using educational robots, etc.

5. The Principle of Practices

5.1. Transparency
Manufacturers and service providers of AI·Robotics should be able to explain the input, dataset for machine learning, internal process, and type and status of motion at the request of stakeholders including police, court, counsel, insurance company and etc., prescribed by law. It should be visualized or explained convincingly in a way that the stakeholders can easily understand. The principle of transparency does not demand to disclose the source codes of artificial intelligence developed by manufacturers or designers. But it requires, in certain case, reasonable explanation about the acquisition path and type of datasets that machine learning and the algorithms used for a series of judgment processes.

5.2. Controllability
Controllability refers to the capability of an operator or user to normalize or stop the system immediately when the AI·Robotics have an abnormality. Manufacturers must ensure that the function that allows the user to instantly control or stop is installed in an easily visible position of AI·Robotics. Manufacturer shall inform users of the function of controlling or stopping the operation of AI·Robotics in accordance with their judgment. Manufacturers shall also fully inform service providers as well as users of the functions beforehand. They are expected to be fully aware of it. The service providers should be able to provide controllability related functionality if the AI·Robotics malfunctions or behaves abnormally, and the user should also be informed of the usage of the related functions.

5.3. Accountability
Accountability is about assumption of responsibility and 'answerability' for motions, and decisions of AI·Robotics caused by malfunctions or abnormalities. Designers of AI algorithm and robotics should have responsibility to provide evidence of potential harms to users or service providers. Manufacturers will have to clarify how decisions are made in case of it. Manufacturers should notify service providers of the possibility of unexpected accidents that may occur, and what types of compensation system are prepared to cover the failure. Providers and users need to be careful to use AI·Robotics products and services in accordance with the purpose of production and its instructions.

5.4. Safety
Manufacturers and service providers should put top priority on the safety of users with regards to usage of AI·Robotics. In this light, specific safe standards of AI·Robots should be provided to ensure the benefit everyone. In order to minimize the safety hazard that can arise from the autonomous judgments and actions, each actors should fulfill their ethical responsibility in

technical design, production, supply, use and management stages. They shall notify users the possibility of risk that defective products or services may cause, even after the supply of the products or services.

5.5. Information Security

The use of AI·Robotics needs to avoid unduly infringing on the privacy of individuals or obtaining excessive user data. Manufacturers and service providers should pay attention to the asymmetry of the information and data generated by AI·Robotics. When each individual expresses his/her consent or refusal of collecting data or personal information Human-Robot Interaction (HRI), the intention should be reflected instantly in the robotic system.

6. Elusive Quest for Liability

As robotics become more intelligent, it will become harder to decide who is responsible if they injure someone both with intention or by accident. The ever-growing number of robotics may bring an increased risk of malfunction of robots, but it can arise from a variety of causes like error in algorithm design, data bias, control panel problems, or mechanical failure. Preventing malfunction only through ethical design would be the most daunting task of the robotics designer. For this reason, the Charter has not directly addressed "who will be held responsible in the event of misbehavior or malfunction?" It would be an *ex-post* scenarios, rather than *ex-ante* one that can be achieved through inherently safe design measures for intelligence robots.

The discussion of liability of robot, one of the most controversial legal conundrums, may fall in front of the threshold of law. Thus, the effective adjudication of the disputes involving the liability of robotics would require a legal analysis and the proof of formal causation rather than an ethical code *per se*. The matter can be done through product liability insurance or strict product liability. Perhaps lawmakers or court may look to alternative models of liability. Thus, liability in robotics seems to need time to reach public consensus, while the standpoints of manufacturers and consumers may vary.

7. Conclusion

This paper provides an overview of the recently proposed Robot Ethics Charter with account of its values and core principles.

Acknowledgments

[1] Turing, Alan (1950), "Computing Machinery and Intelligence", *Mind*, LIX (236): 433–460,

[2] Boer Deng (2015) "Machine ethics: The robot's dilemma" *Nature* 523, 24–26, doi:10.1038/523024a

[3] Establishing a Korean Robot Ethics Charter by Robot Division, Ministry of Commerce, Industry and Energy, (April, 2007) www.roboethics.org/icra2007/contributions/slides/Shim_icra%2007_ppt.pdf

[4] "Military work threatens science and security" *Nature* 556, 273 (2018)

[5] Intelligent Robots Development And Distribution Promotion Act of Korea. Article 5.2 (Revised June. 12, 2018)