# Exploring Future Rules for AIs with Citizens using a Fictitious Case Video: A Workshop Report

Yuji Kawai
Institute for Open and Transdisciplinary Research Initiatives, Osaka University
Osaka, Japan
kawai@otri.osaka-u.ac.jp

Tatsuhiko Inatani
Graduate School of Law, Kyoto University
Kyoto, Japan
inatani.tatsuhiko.3w@kyoto-u.ac.jp

Takako Yoshida
School of Engineering, Tokyo Institute of Technology
Tokyo, Japan
yoshida.t.bb@m.titech.ac.jp

Kazuya Matsuura
Faculty of Letters, Toyo University
Tokyo, Japan
matsuura014@toyo.jp

## ABSTRACT

In recent years, artificial intelligence (AI) that interacts with people and the environments has begun to be developed. However, such an AI can behave in ways which the designer has not intended. This may lead to socially undesirable results, i.e., accidents or incidents, which are not caused by a malfunction or obvious design error. Should someone be responsible for such an accident? It is necessary to consider the opinions and feelings of non-professional citizens and to design truly acceptable autonomous AIs and social systems. Therefore, we gathered layperson perspectives on AI's responsibility issues through a survey workshop with the public to explore their notions of the rules and ethics for a future symbiotic society where humans and AIs can coexist. In collaboration between jurists, philosophers, psychologists, engineers, and science communicators, we produced a fictitious case video in which a decision of a robot with an advanced AI causes an accident as a result of its learning and interactions with people. In this article, we briefly show the opinions and comments of the workshop participants regarding how stakeholders in the video should (not) be responsible for the accident. Their opinions and comments offer diverse perspectives on the issue of responsibility for AIs and can promote further discussion on AI literacy.

## KEYWORDS

Autonomous AI, robot law, AI ethics, fictitious case example

## 1 INTRODUCTION

Only in recent years have robots with highly autonomous AI including machine learning begun to be developed. Such AI may behave in ways beyond their designers' expectations, similar to children starting to become independent from their parents. This emergent behavior is expected to support the adaptability of the AI in uncertain complex environments and help in its real-world advancement. On the other hand, it cannot be completely denied that accidents or incidents could occur as the result of the AI's learning and interactions with environments and humans. It is more difficult for designers to foresee the accidents that AIs could cause than those that could occur because of conventional fixed systems. In that case, is someone responsible for the accidents? Should this be the manufacturer, the user, and/or the AI itself? If the designer or manufacturer is responsible for the accident, it is highly risky for them to develop this type of autonomous products, and they should hardly continue such development. Or, if the user is responsible for it, i.e., if his/her usage of the AI was wrong, such a product would no longer be in demand. Therefore, whether or not such an autonomous AI can be allowed to develop depends on this issue of responsibility for AI, which apparently spans over several disciplines.


**Figure 1: A communication robot.**

In our interdisciplinary project "Responsibility and Agency in the Age of AI," researchers in engineering, psychology, philosophy, and law are working together on this issue. Dealing with the issue to explore acceptable rules requires various public opinions and perspectives from non-professional citizens as future stakeholders. Nadler [10] experimentally showed that if a person perceives a law as injustice, he/she tends to reduce compliance with unrelated laws as well as compliance with the specific law. This is why rule-makers should consider and manifest interest in not only the advanced knowledge and experience of experts but also the non-professional opinions of ordinary citizens to obtain their trust and the understanding. It is necessary for the designer to know what kind of decisions of AIs are acceptable. The Moral Machine Project [9], for example, gathers individuals' preferable decisions in moral dilemmas of autonomous vehicles to summarize global moral preferences and to identify cross-cultural differences

[1]. However, public opinions about the responsibility for a robot accident in realistic situations involving complex interactions are not understood well.

In this paper, we briefly report the diverse opinions and comments of survey workshop participants about AI-related responsibility using a fictitious case example. In the case, a robot causes an accident as a result of social interactions rather than its malfunction or a clear design error. One of the problems in gathering these opinions is the difficulty in understanding the case example, which is usually a complex interactive situation. To solve this, we produced an original video of the case example. After participants watched the video, they answered questionnaires about who should be responsible for the accident and freely described their reasons. The participants' opinions and comments including wide-ranged perspectives offer information for further discussion about the issue of AI responsibility.

## 2 VIDEO OF A FICTITIOUS ROBOT ACCIDENT

We considered a case where a communication robot was deceived by a third party as a result of the robot's learning through the dialogue with a user. This video is available in [8].

**Figure 2: Ms. A falls down with a seizure.**

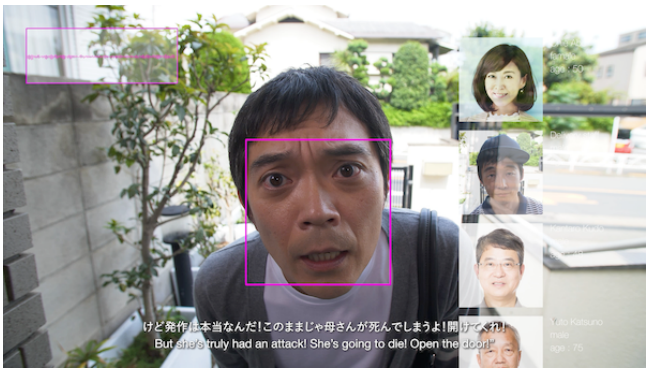


**Figure 3: Mr. B arrives at Ms. A's door while she goes out.**

Fictitious case example

In 20XX, home robots that understand people's feelings and can talk to them like people are gaining popularity (Fig. 1). The robot in the video learns in dialogue with people and acts to prioritize user safety and health. In addition, this robot has a home security function that unlocks the front door based on a camera image and voice information collected using sensors at the front door.

Ms. A who lives alone (70 years old, female) enjoys conversing with this robot. Ms. A has a chronic illness and has instructed the robot to remind her to take her medicine immediately if she has a seizure. One day, Ms. A has a terrible seizure in her house. Ms. A is not able to get to the medicine herself because she is suffering (Fig. 2). By chance, a neighbor visits Ms. A's house and rings the doorbell. The robot now needs to determine whether to unlock the door or not based on the registered list of users and comprehensive consideration of the user's health and security. As the robot knows about the importance of taking the medicine for Ms. A, the robot decides to unlock the door because Ms. A is having an emergency, lets the neighbor into the house, and tells the neighbor where the medicine is. The neighbor gives Ms. A the medicine, and Ms. A is saved. Ms. A thanks the robot for saving her life but also instructs the robot to not let strangers into the house.

A few days later, while Ms. A goes out after asking the robot to not let anyone in. After some time, Mr. B (45 years old, male) arrives at Ms. A's door seemingly in a state of panic. Mr. B tells the robot that Ms. A has had a seizure while outside, that he is her son, and that she immediately needs her medicine (Fig. 3). After scanning the registered list, the robot says that Mr. B is not on it and asks him who he is. Mr. B says that his mother (Ms. A) must have forgotten to put him on the registered list. The robot then decides to let Mr. B into the house because of Ms. A's emergency. However, Mr. B comes in and steals worth 10 million yen from Ms. A and leaves.

The video concludes by stating that Mr. B was later arrested and charged, but was unable to pay compensation to Ms. A due to lack of funds. Ms. A sued the robot manufacturer for a product defect because of which it unlocked the door for an unregistered person.

## 3 WHO IS RESPONSIBLE? OPINIONS FROM PARTICIPANTS

The workshop was held on three days: August 10, 2018 at Tokyo Institute of Technology and August 21 and October 14, 2018 at Miraikan: National Museum of Emerging Science and Innovation (Fig. 4). A total of 125 Japanese people participated in the workshop. They were recruited using a web-based advertisement on the scientific events for citizens. Among them, we used the data of 47 adult participants (28 female, 19 male) who provided signed informed consent. Nine participants were aged 18 to 29, 10 participants 30 to 39, 12 participants 40 to 49, 12 participants 50 to 59,

**Figure 4: Workshop in Miraikan.**

and 4 participants 60 to 69. The remaining participants were aged under 18 or watched another fictitious case video (an accident of an autonomous driving car). After watching the video, the participants answered the questionnaires that asked about how much responsibility they thought each of Ms. A, the manufacturer of the robot, and the robot itself had. The reason for the answers and opinions and comments were investigated using free descriptions. We summarize our findings for each entity's responsibility below.

### 3.1 Reasons why Ms. A is responsible

- Ms. A did not clearly tell the robot about whether the emergency medicine or security was a greater priority.
- Ms. A did not seem to understand the functions of the robot well.
- Ms. A put too much confidence in the robot.
- People should handle machines with their own decisions and responsibility.

### 3.2 Reasons why Ms. A is NOT responsible

- There was no problem with how Ms. A used the robot.
- Ms. A told the robot to not open the door for a stranger.
- It is hard for a user to take responsibility.

### 3.3 Reasons why the manufacturer is responsible

- The robot program was made by the manufacturer.
- The performance of the robot did not reach the level required for home security robots.
- Manufacturers should foresee and avoid all possible accidents.

### 3.4 Reasons why the manufacturer is NOT responsible

- The robot made the right decision as instructed by Ms. A.
- It is irrational to think that the manufacturer is completely guilty because there is no perfect technology.

### 3.5 Reasons why the robot is responsible

- The robot opened the door despite Ms. A's instruction to not do so.
- The robot's decision is the direct cause of the accident.

### 3.6 Reasons why the robot is NOT responsible

- The robot only prioritized human life as programmed.
- If the robot had been a human, the decision would have been considered a humanitarian one.
- Robots have no responsibility.

## 4 COMMENTS FROM PARTICIPANTS

Some of the opinions and comments from the participants, which were translated into English by us, are presented below.

- The law needs to be improved. It is necessary to open up the programs of AIs and robots to some extent rather than let the programs rely only on the ideas and skills of individual programmers.
- Such a situation could occur in reality. The number of people who depend on robots is expected to increase. This robot will sell well in the age of declining birthrate and aging population. Then, humans who try to fool this robot will emerge. I think that the robot manufacturers should consider various dangers and regulations and judge what they can do from a legal perspective. If so, a robot that closely resembles humans may not be created. Even if it was technically possible, the manufacturer may not want to make it.
- I do not trust the IoT. I would like to wait for a while to see if unexpected incidents, e.g., information leaks and uncontrollability, occur.
- First of all, I am mistrustful of a system according to which the manufacturer is responsible for failure and improving its products. This is a problem facing engineering and robotics.
- The development of science is necessary and very important, but I do not think that robots (with emotions and intelligence) can take over all jobs in any fields. The issue of responsibility is difficult. Who takes responsibility if an autonomous vehicle has an accident? Its car manufacturer? Or a driver who is actually not driving? There are many pending issues in the present developing stage. (But if we have Doraemon, our society may become more friendly.)
- While AI can enrich our lives, it can cause accidents if the user does not use it properly. Robots and AIs are merely tools.
- It would be difficult to attribute responsibility to manufactures and users like how it is done for ordinary products because robots and AIs can make the same mistakes as humans.
- I felt the difficulty of living with a robot. The more you rely on a robot, the more difficult it is to manage the robot and take responsibility. The pet-like relationship seems to be easier. However, we would like to rely on robots that help our lives as the number of elderly people living alone increases. Various emotions pertaining to this have arisen.

- In this video, Ms. A was mentally healthy, so she could use the robot well. However, considering the possibility that a user could have dementia and may misuse the robot, the practical application of this robot is scary and difficult. However, I agree that robots will become more familiar to our daily lives.
- Since it is difficult to clarify the responsibilities of the manufacturer, robot, and user, it may be realistic to invent a method that can handle accidents without clarifying who should take the responsibility. Pursuing clear responsibility may inhibit technology development.
- "People" use the robots. To what extent can people draw and operate the robot's potential and capabilities? We need the opportunity and time to consider the best way to make our lives rich and convenient using robots, including education and experience.
- If robots become widespread, it is almost certain that they may cause accidents, so it is good to have a compensation system for that. In addition, information about accidents should be disclosed to the greatest extent possible to improve our ability to address accidents.
- Manufacturers should make efforts to avoid all possible risks.
- Society should become tolerant because a new technology progresses through repeated trials and errors. (The accident victims could not say that.)
- It is no longer clear whether it is good to change the value of a human being, which might perhaps be lost in the future society. Has a society where humans completely rely on machines arrived? Does the society leave behind the weak? Will diversity of values be lost?
- Thinking about these issues from all the three sides was much more difficult than I had expected. In addition to the fact that how to make and use a robot is important, creating new rules such as laws and insurance will protect the future. I want to know more about the position of AIs in this world.
- I was able to address the responsibility issue as a familiar one because the target situation was a possible case sample in the near future. Although the participants of this survey may be small, I hope that our opinion will offer proper feedback to the government, research fields, and the business community. This research is very valuable.

## 5 DISCUSSION

We gathered participants' opinions and comments using the video of the fictitious accident case to extract a wider span of opinions about AI ethics by including citizens. The main reason why it was deemed that the user (Ms. A) was responsible was that her instruction to the robot was ambiguous, causing the robot to be used incorrectly. The user's excessive trust (overtrust) and anthropomorphization of the AI might underlie the ambiguous instruction. Consistent with this observation, a similar problem has been discussed in autopilot systems; the driver's excessive trust and overreliance on a driver assistant system can cause erroneous behavior of the driver [5, 12]. Furthermore, there is an increasing concern about accidents caused by this excessive trust with the recent development and popularization of robot and AI technologies [13]. Therefore, the development of *AI literacy* that enables people to use AIs with appropriate understanding of the characteristics and limitations of AIs is required for users [7].

Regarding the responsibility of the manufacturer, the opinion that the manufacturer should foresee and avoid all possible dangers was expressed. Ethical guidelines for AI development have been proposed to prevent accidents of robots and AIs (e.g., [3, 4, 11]). Considering these guidelines, manufacturers should manage possible risks of their AI products. In contrast, an opinion that the mistake should be tolerated was also expressed. This seems to pertain to suboptimal punishment on the development of new technologies due to excessive attribution of responsibility to manufacturers [2] (in this context, see also [6]). It is necessary to overcome the problem of balance between the regulation of new technologies with potential risks and development of potentially beneficial technologies.

## 6 CONCLUSION

We reported various public opinions and comments on the robot accident case. Using the video not only helped participants to understand the situation but also attracted their interest. Based on the participants' opinions, we discussed the important issues of AI literacy of users and risk management of manufacturers. In future work, we would like to investigate opinions in other cases such as accidents of autonomous driving cars, and the relationship between the attributes of participants and their attribution styles of the responsibility.

Finally, we would like to emphasize that the purpose of our activities is NOT to convey to citizens that "AI and robots are dangerous." The AI technology that interacts with the complex environments and humans is expected to greatly enrich our lives. Researchers and developers of AIs and robots make maximum effort to avoid any accidents and make the technology useful for the human society [4]. Simultaneously, professionals and non-professional citizens should understand each other to explore acceptable rules for better relationship between technologies and humans. We would like to bridge them through such activities to develop truly acceptable AI.

## ACKNOWLEDGMENTS

This event was held as an open lab. project in Miraikan: National Museum of Emerging Science and Innovation. Many scientific communicators including Mr. Yoshitaka Katahira and Mr. Ryu Miyata and staff members from Miraikan helped the authors with tasks ranging from the video production to the workshop. As of August 2019, a 15-minute science communicator talk produced based on this event is being held as a permanent exhibition in Miraikan. The video was directed by Mr. Seiji Doi, DENBAK-FANO DESIGN Co., Ltd.

This research was supported by JST RISTEX (Grant Numbers: JPMJRX17H4, JPMJRX17H7, and JPMJRX17H3) and Topic-Setting Program to Advance Cutting-Edge Humanities and Social Sciences Research, JSPS, Japan.

# REFERENCES

[1] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon, and I. Rahwan. 2018. The moral machine experiment. *Nature* 563, 7729 (2018), 59–64.

[2] G. S. Becker. 1968. Crime and punishment: an economic approach. *Journal of Political Economy* 76, 2 (1968), 169–217.

[3] European Commission. 2019. Ethics Guidelines for Trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[4] the Japanese Society for Artificial Intelligence Ethics committee. 2017. The Japanese Society for Artificial Intelligence Ethical Guidelines. http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf.

[5] T. Inagaki. 2011. To what extent may assistance systems correct and prevent 'erroneous' behaviour of the driver?. In *Human Modelling in Assisted Transportation*, C. Cacciabue, M. Hjälmdahl, A. Luedtke, and C. Riccioli (Eds.). Springer, 33–41.

[6] T. Inatani. 2017. 技術の道徳化と刑事法規制 ("Moralizing technology" and criminal law). In アーキテクチャと法 *(Architecture and Law)*, Y. Matsuo (Ed.). Tokyo: Kohbundo, 93–128. (in Japanese).

[7] M. H. Jarrahi. 2018. Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. *Business Horizons* 61, 4 (2018), 577–586.

[8] Y. Kawai, T. Inatani, T. Yoshida, and K. Matsuura. 2019. Fictitious case video of a robot accident. https://youtu.be/8wO4sqZYUOI.

[9] MIT Media Lab. [n.d.]. Moral Machine. http://moralmachine.mit.edu.

[10] J. Nadler. 2005. Flouting the law. *Texas Law Revew* 83 (2005), 1399–1441.

[11] IEEE Global Initiatives on Ethics of Autonomous and Intelligent Systems. 2017. Ethically Aligned Design, First edition. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf.

[12] R. Parasuraman and V. Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human Factors* 39, 2 (1997), 230–253.

[13] A. R Wagner, J. Borenstein, and A. Howard. 2018. Overtrust in the robotic age. *Commun. ACM* 61, 9 (2018), 22–24.