ORIGINAL PAPER

Moral Awareness of College Students Regarding Artificial Intelligence

Nader Ghotbi[1] · **Manh Tung Ho[1,2]**

© National University of Singapore and Springer Nature Singapore Pte Ltd. 2021

Abstract

To evaluate the moral awareness of college students regarding artificial intelligence (AI) systems, we have examined 467 surveys collected from 152 Japanese and 315 non-Japanese students in an international university in Japan. The students were asked to choose a most significant moral problem of AI applications in the future from a list of ten ethical issues and to write an essay about it. The results show that most of the students (n=269, 58%) considered unemployment to be the major ethical issue related to AI. The second largest group of students (n=54, 12%) was concerned with ethical issues related to emotional AI, including the impact of AI on human behavior and emotion and robots' rights and emotions. A relatively small number of students referred to the risk of social control by AI (6%), AI discrimination (6%), increasing inequality (5%), loss of privacy (4%), AI mistakes (3%), malicious AI (3%), and AI security breaches (3%). Calculation of the z score for two population proportions shows that Japanese students were much less concerned about AI control of society (-3.1276, $p < 0.01$) than non-Japanese students, but more concerned about discrimination (2.2757, $p < 0.05$). Female students were less concerned about unemployment (-2.6108, $p < 0.01$) than males, but more concerned about discrimination (2.4333, $p < 0.05$). The study concludes that the moral awareness of college students regarding AI technologies is quite limited and recommends including the ethics of AI in the curriculum.

✉ Nader Ghotbi
nader@apu.ac.jp

Manh Tung Ho
tung.homanh@phenikaa-uni.edu.vn

[1] Ritsumeikan Asia Pacific University, Oita, Beppu, Japan

[2] Centre for Interdisciplinary Social Research, Phenikaa University, Yen Nghia, Ha Dong, Hanoi, Vietnam

Introduction

New technologies such as gene editing, nanotechnology, and artificial intelligence (AI) may lead to new applications thereof with ethical implications that have not been examined before. AI is a rapidly developing technology with wide applications in and by the society which leads to the concern whether people are aware of the associated nuances of these applications. While experts in AI may discuss and debate the normative policies and regulations surrounding the use of AI, the civic society may fall behind because of an asymmetry in technical information and awareness about the possibilities that lie ahead. Therefore, we designed this study to examine the level of awareness of the educated youth regarding the ethical concerns associated with the use of AI. Moral issues in artificial intelligence (AI) are associated in general with the ethical principles of benefit vs. harm, justice and fairness, moral agency and motivation, and transparency (Boddington 2017). Benefits of AI result from the way it generates new insights and knowledge by examining huge amounts of data, optimizing processes, and increasing efficiency and productivity which also lead to economic profit and wealth in society, and may contribute to environmental sustainability (Stahl et al. 2021). However, these benefits produce certain risks that need to be understood and controlled (Yigitcanlar et al. 2020; Stahl et al. 2021). Floridi et al. (2018) refer to the risks of underusing (opportunity costs) and mis- or overusing AI which include "devaluing human skills, removing human responsibility, reducing human control, and eroding human self-determination." Bossman (2016) listed nine potential ethical issues in AI including unemployment, inequality, impact on human behavior and interaction, committing mistakes, racism and bias, security threats, malicious acts, loss of human control, and rights for robots. A more recent review by Stahl et al. (2021) refers to the loss of transparency, accountability, privacy, and individual freedom; potential for misuse, mistakes, bias, and discrimination; impacts on vulnerable groups, employment, security, democracy, and human nature; and the rise of autonomous machines and, possibly, machine consciousness. This section presents a short review of ten specific ethical issues in AI.

Unemployment

Wilkinson and Marmot (2003) referred to the social, psychological, and health benefits of work; losing such benefits because of AI can be considered as ethical harm. Mainichi (2016) reported that insurance companies transferred the job of their human staff to IBM Japan Ltd.'s Watson, an AI system with "cognitive technology that can think like a human." Therefore, some job losses may already have happened or be underway because of substitution by AI systems. However, any new technology since the industrial revolution has caused changes in the employment market, and the overall impact of AI on employment is complicated with certain jobs more at risk, such as those that involve highly repetitive or structured actions in a predictable setting (Boddington 2017). On the other hand, jobs requiring interaction with

people and using social intelligence, creativity, and clever solutions and working in an unpredictable environment are less at risk of being filled with AI (Tegmark 2017). Walsh (2018) has noted that AI experts consider the impact of AI on employment to happen decades later than what non-experts assume, thus allowing much time for people to adapt to the changes in the employment market.

Emotional AI

This includes, first, the sensing and perception of human emotions by AI technologies and, second, the impact of robots on human emotions as well as the robots' emotions and rights. The former is an issue better discussed under "privacy", while the latter has been a hot topic in Japan. Ozawa (2021) reports on the successful sale of many robots in Japan that appeal to the emotions of Japanese customers, such as the "chatty" *Charlie* (from *Yamaha*), the "humanoid" *Robohon* (*Sharp*), the "dog" robot *Aibo* (*Sony*), the "friendly" *Pepper* (*Softbank*), the "pet-like" *Qoobo* (*Yukai Engineering*), and the "heartwarming" *Lovot* (*Groove X*). This level of emotional interest in robots in Japan has been explained by the prevalent view in Japanese culture that objects too can have a soul, and that Japanese customers are interested in the "character" of their robot not just its mechanical function. One may argue that machine emotions are fake, while those of humans are real. However, one may respond that human emotions in a social context are also commonly faked; the concept of emotional intelligence (EQ) is the ability to read emotions expressed by others and respond in "appropriate" ways (Fiori and Vesely-Maillefer 2018). In other words, faked emotions are as practical as real ones. Moreover, the question of whether intelligent robots in the future should be granted some form of rights for feeling emotions depends crucially on whether they will be conscious and can subjectively suffer from sadness or feel joy and other emotions (Tegmark 2017).

AI Mistakes

The issue of AI safety is a good example to demonstrate the complexity of risk vs. benefit assessment. Leslie (2019) has produced a guideline to help identify the potential harms caused by AI systems and suggested practical measures to anticipate and prevent them through responsible innovation and governance. He recommends that the implementation of AI systems be human-centered with emphasis on evidence-based reasoning, situational awareness, and moral justifiability. Mistakes committed by autonomous cars have been publicized, and the problems of non-predictability and machine autonomy appear to diminish the level of trust of the public to the use of AI (Alaieri and Vellino 2016). Responsibility over AI mistakes is particularly significant in healthcare. Examining the use of AI in robotic surgery, O'Sullivan et al. (2019) classify responsibility into accountability, liability, and culpability; the supervising human surgeon plays the key role in protecting the patients undergoing operations from potential mishaps.

AI Control of Society

Western media and literature commonly depict frightening scenarios about misbehavior of autonomous robots and their use for social control (Liang and Lee 2017; Zhang and Dafoe 2019). However, Japanese society demonstrates a positive image of robots and their character with robots having been accepted as partners of humans both at home and at the workplace; the Japanese government promotes the use of "social" robots to ameliorate the problems of an aging society, particularly the dwindling number of the young manpower to replace those retiring from the workforce (Wagner 2009). These examples contrast with the use of AI for the surveillance and control of the masses of people elsewhere (Polyakova and Meserole 2019). For example, China's "social credit system" has extended the use of AI technologies for political control of the society in a coercive system based on modern surveillance techniques (Hoffman 2018; Curran and Smart 2021). In Russia, several tools are used by those in power for administrative control over the digital social space as a "modern policy-management alternative" (Mikhaylenok and Malysheva 2019). An extensive literature review by Hagerty and Rubinov (2019) has demonstrated several patterns of AI usage for social control at a global scale and suggests that low- and middle-income countries are more vulnerable to the negative social impacts of AI while they're also less likely to benefit from it.

Increasing Inequality

Unless a fraction of the AI-created wealth is redistributed to make everyone better off, inequality may increase significantly (Korinek and Stiglitz 2019). AI can increase the share of capital in the economy as compared with that of labor and increase the income for certain groups of highly educated/skilled people while making it harder for others to earn a livelihood (Agrawal et al. 2019). Therefore, the economic gap in the society may become wider unless new policies are put in place, such as taxation of capital and a basic universal income.

Discrimination and Bias

Ntoutsi et al. (2020) have explained in detail how problems in gathering and processing of data may result in biased AI decisions over human characteristics such as race and gender; they provide references in literature to many instances of racial and gender discrimination. They recognize that human societies suffer from deeply embedded biases that cannot be eliminated with only technical solutions; multidisciplinary approaches including social and legal remedies are also needed to avoid prejudice.

Privacy

The use of AI by governments for surveillance of global citizens at home and in public spaces through massive data collection creates serious ethical concerns but has commonly been justified by the need for security (Bartneck et al. 2021). Private companies and businesses also intrude into the privacy of citizens to improve advertising, marketing, and sales with the justification that it is the basis for better performance of AI systems. The use of psychological and emotional AI techniques helps manipulate and exploit human users and erodes their right to privacy.

Malicious AI

AI can be used to harm people physically. King et al. (2020) have provided a systematic, interdisciplinary analysis of the literature about the foreseeable threats of AI crime, which is the use of AI technologies to facilitate criminal acts such as automated fraud targeting social media users, and the manipulation of simulated financial markets. The use of AI for psychological warfare by synthesizing fake human images called deepfakes has been explained by Pantserev (2020). One can only expect that AI with malicious intent will be used more widely in the future by organized crime. This brings us to the next ethical issue which is about the vulnerability of AI systems to security breaches.

Security Risks

Yigitcanlar et al. (2020) have referred to several cyberattacks against the AI infrastructure of smart cities which caused massive dysfunction in their communication systems (telephone and email), law enforcement, waste, energy, and payment systems. The city councils often had to pay ransoms to those who had breached the security system or employ external cybersecurity and consulting companies to deal with the situation. It appears as the implementation of AI systems creates new opportunities for criminals to use its vulnerabilities and cause wider damage to communities.

As it was briefly explained in this section, there are a variety of moral issues that need to be considered before implementation of AI-based technologies. These issues are rapidly in progress and appear to be serious, but understanding their magnitude requires technical knowledge about the way AI systems have been used so far. Therefore, we decided to conduct a survey in an international university in Japan to examine the moral awareness of college students (Japanese vs. non-Japanese) regarding AI and to understand how much the education system has contributed to their understanding of moral issues surrounding AI. The study was done at Ritsumeikan Asia Pacific University (APU) which has about 5700 students with an almost 54 to 46% mix of Japanese and non-Japanese students. Considering that Japan is a pioneer in the development of robotics technology and its shrinking population and workforce

would require a wider use of AI technologies to substitute for manpower, this environment provided an opportunity to examine whether there is a meaningful difference between responses by Japanese students vs. those from other countries. The current study is focused on ethical issues of AI after we published a study using sentiment analysis on a smaller sample of students (Ghotbi et al. 2021).

Research Methods

In large classrooms of college students enrolled in a general course of bioethics at foundation level, we asked students for their cooperation in a study about the ethics of AI. They were presented with the survey question and requested to upload their answers to a digital repository. Out of 478 students, 11 students chose not to cooperate with the survey so that we could collect 467 (97%) responses after a month. The respondents included 152 (33%) Japanese and 315 (67%) non-Japanese students mainly from China, Korea, Vietnam, Thailand, Indonesia, Uzbekistan, Mongolia, India, Bangladesh, and a few from other nationalities. The survey question was as follows:

> Artificial Intelligence (AI) has the potential to cause a number of ethical issues in the future. *Which* one of the following do you consider as the most important and explain *why*: AI control of society, AI discrimination or bias, AI impact on human behavior and emotions, AI increasing inequality, AI mistakes, AI security risks, loss of privacy, malicious AI, robots' rights and emotions, unemployment.

Our argument for requiring an explanation for the selected ethical issue was to incite reflection of the students and lower the chance that they would choose an item on impulse. There were a number of research questions to be investigated through this survey. For instance, are there differences in the moral sensitivity of the educated youth towards the possible misuse of AI technologies, based on gender and nationality? Do many of them worry about the loss of privacy as AI applications are increasingly used to swift through personal data for various purposes? Are they more concerned about AI being used by authoritarian governments trying to control the society through AI, or the possibility that in the future they may have to compete with AI when looking for employment? The responses to the survey were collected from the digital repository and examined one by one to identify the selected ethical issue and look for correlations with basic demographic characteristics.

Findings and Results

Among the 467 students who responded to the survey question, there were 152 (33%) Japanese and 315 (67%) non-Japanese students; 236 (50.5%) were female and 231 (49.5%) were male; 52 (11%) were freshmen (1st year), 134 (29%) sophomores (2nd year), 164 (35%) junior (3rd year), and 112 (24%) senior (4th year); and 5 students (1%) were at their 5th or 6th year in college. As for age, all respondents were

Table 1 Students' response to which ethical issue would be the most important in AI in the future

Ethical issue	Number of students by nationality and gender					
	Japanese/non-Japanese		Female/male		Sum/%	
Increasing unemployment	93	176	122	147	269	57.6%
AI impact on human behavior and emotions	14	23	23	14	37	7.9%
AI control of the society	2	28	14	16	30	6.4%
AI discrimination, racism, or bias	15	14	21	8	29	6.2%
Increasing inequality	6	19	17	8	25	5.4%
Loss of privacy	5	13	10	8	18	3.9%
Robots' rights and emotions	3	14	11	6	17	3.6%
AI mistakes	7	9	6	10	16	3.4%
Malicious AI	3	11	8	6	14	3.0%
Security risks	4	8	4	8	12	2.6%
Total	*152*	*315*	*236*	*231*	*467*	*100%*

between 19 and 24 years old. The written essays varied in length, but were mostly around 1100 words, typed in about two pages of single-spaced common font size 12. They include various arguments and reasoning based on material they had seen, read, or heard about, to explain why the topic they had chosen was a significant moral issue of AI. The results of students' choice of which ethical issue would be the most important are shown in Table 1.

As seen in Table 1, among the ten ethical issues possibly arising from AI, the most common concern of college students in this sample was over unemployment (no = 269, ~58%). The next most common concern was over emotional AI issues, including the AI impact on human behavior and emotions, and robots' rights and emotions (no = 54, ~12%). Unemployment and emotional AI together represented the concerns of 70% of all students, with the remaining 30% going to AI control of society, AI discrimination, increasing inequality, loss of privacy, AI mistakes, malicious AI, and security risks. Next, we checked for statistical significance in the size of differences observed between the proportion of students selecting a particular moral concern and their gender as well as Japanese (vs. non-Japanese) nationality. The *z* score formula for comparing two proportions is:

$$z = \frac{n(\rho 1 - \rho 2) - 0}{\sqrt{\rho(1 - \rho)(^1/_{n1} + ^1/_{n2})}}$$

Checking for differences in the proportion of Japanese vs. non-Japanese students selecting *AI control of society*, the value of z was − 3.1276 with the result being significant at p < 0.01. Checking for differences in the proportion of Japanese vs. non-Japanese students selecting *discrimination*, the value of z was 2.2757 with the result being significant at p < 0.05. Checking for differences in the proportion of female vs. male students selecting *unemployment*, the value of z was − 2.6108 with the result

being significant at p < 0.01. Checking for differences in the proportion of female vs. male students selecting *discrimination*, the value of z was 2.4333 with the result being significant at p < 0.05. All other differences seen in the collected data were statistically insignificant at p < 0.05. These results support the alternative hypotheses that fewer Japanese students than non-Japanese students are concerned about "AI control of society", but more of them are concerned about "discrimination" which was also a more common concern among female students, while they were less concerned over "unemployment" than male students.

Discussion of the Results

The current study builds up on an earlier work we did using textual sentiment analysis and focusing on the emotional reactions of survey respondents to the question of AI (*Citation Blinded*). This work is distinct in several ways: First, we have reviewed the ethical aspects of AI with a reference to the latest articles published, adding privacy as an important moral issue. Secondly, we have used quantitative methods to provide evidence for two novel findings. One is a significant lack of concern over the risk of social control among Japanese respondents. In literature review, we found only one paper discussing this aspect of Japanese social ethics (Wagner 2009). The other finding is that Japanese and female respondents were significantly more concerned, than the non-Japanese, with discrimination which is another evidence pointing to the prevalent problem of gender discrimination in Japan.

The most common concern of college students in this study, whether Japanese or non-Japanese, male or female, was over possible unemployment related to the use of AI. However, this concern was significantly more common among males. There are several possible reasons for this difference. It could be that males in Asian culture consider employment as more vital for their livelihood, or that more females were concerned with other ethical issues of AI such as discrimination which is a particular issue at work (Hara 2018; Peillex et al. 2019). As explained briefly in the introduction section, although unemployment is a major concern for the future, AI experts believe it will take a few more decades to see a large impact on employment (Walsh 2018), while certain uses of AI have already caused serious moral issues. Therefore, the significantly higher level of interest in other moral problems among female students may reflect more awareness about the ongoing social issues, especially gender discrimination. It also raises the possibility that their experiences of gender discrimination may have been extended to their expression of worries over the risks of AI in the future.

The second most common concern of the students was over emotional AI, including the AI impact on human emotions and the rights and emotions of robots themselves. But is concern over emotions of robots and their rights a genuine ethical concern? Will intelligent robots in the future be conscious and capable of feeling sadness, joy, and other emotions? In a bioethical perspective, the necessary conditions for having consciousness include being a living thing, as opposed to the non-living such as rocks, and being alive and arousable, as opposed to being dead. Is it possible that AI systems may gain consciousness in the future, while so far

only living things are presumed to be conscious? We say "presumed" because consciousness is a "subjective" sense of awareness and a perception of the environment around the "subject". Humans can share their subjective sense of awareness and consciousness with other humans and thus learn about its character, but how can they be sure about the level of consciousness in animals, plants, and simpler living organisms? The science of biology may help overcome this hurdle to some extent; consciousness depends on a bodily system to sense and perceive the environment outside the subject which is mainly chemical in plants and simpler forms of life and "neurochemical" in animals including humans. The term neurochemical implies that neural activity by nerve cells and chemical activity by neurotransmitters are both involved. Humans appear to have a higher level of consciousness because of their more sophisticated nervous system especially their brain cortex which has developed high degrees of self-awareness that relies on personal memories, personal traits (learned ways of behavior), personal wants and choices, and recognizing personal differences relative to other humans. These together enable a particular focus on self that one may describe as self-awareness. Can a "non-living" AI system gain self-awareness? For consciousness to exist, how essential is the element of life, versus a complex human-like brain that AI is predicted to possess in the future? Does AI need to feel emotions subjectively, like a living human, to have consciousness? If so, is it possible that a cyborg could be the intermediary on the path to building artificial consciousness? We have no definitive answer to these questions, and this would be a good reason to not consider robots' rights and emotions a primary concern.

The possible impact of robots on human emotions is a topic under investigation. Researchers have so far found that human recognition of robotic emotions does not happen automatically and requires more than physical mimicry but may happen when the human user expects emotional expression by robots especially if the robot communicates using both voice and motion (Winkle and Bremner 2017, p. 633). Meanwhile, science fiction films have depicted various scenarios in which future robots express emotions and engage in an emotional relationship with humans (Lorenčík et al. 2013; Schofield 2018). The results of our survey suggest that such films may have had a bigger impact on the responses to the survey than the realistic information produced so far by researchers investigating such a situation. This result is not unexpected because popular culture through media and films may play a bigger role in shaping the views of the youth who mostly have not encountered actual robots.

The number and percentage of students concerned over increasing inequality (5%), loss of privacy (4%), AI mistakes (3%), malicious AI (3%), and AI security breaches (3%) were too small. However, some of these issues are ethically at least as important as employment. The possible use of AI for social control is one of such issues, and we referred to many studies that have documented such misuse of AI (Hoffman 2018; Hagerty and Rubinov 2019; Mikhaylenok and Malysheva 2019; Polyakova and Meserole 2019; Curran and Smart 2021). However, Japanese students were significantly less concerned about social control than non-Japanese students. This may be explained by the unique position of Japanese society over robots (Wagner 2009; Hara 2018). They may also know little about the potential of AI misuse in the future.

Increasing inequality was not a top contender among students' concerns, while it should be. O'Neill (2016) explains how "ill-conceived" mathematical algorithms aggravate inequality in a data driven financial system. One reason may be a lack of sensitivity towards economic inequality among our respondents, which has been reported by Rodriguez-Bailon et al. (2017) in the form of economic system justification and social dominance orientation beliefs.

Overall, the review of the literature at the opening section identified some significant ethical issues associated with the use of AI that have already occurred. However, the survey results demonstrate that most students were either not aware of those problems or were significantly more concerned that AI may affect their employment, even though a significant impact on the job market may take two decades or so. For example, the number of students mentioning the loss of privacy was too small even though it is a very common issue. A harsh critic might even suggest that these responses reflect the everyday concerns of the students rather than their awareness of the types of ethical issues that AI applications have already caused or are expected to cause in the near future.

Conclusion

This study of a large sample (n=467) of college students in a multicultural university with many international as well as Japanese students demonstrated that among them the most common ethical concern associated with AI technologies was over unemployment. However, the review of literature about the moral issues related to the misuse of AI, for instance, the loss of privacy with the ubiquitous use of real-time surveillance, gender- and race-biased systems, the potential for mistakes without due consideration of responsibility, increasing inequality in and among societies, malicious use of AI, and increased vulnerability to security breaches demonstrated that moral issues surrounding the use of AI are complex, while the general level of awareness of the majority of college students was quite limited and mainly focused on the possible loss of employment. Comparing the responses by Japanese students to those by the non-Japanese, statistically significant differences were found in two areas: AI use for social control and discrimination by AI. Japanese students demonstrated a relative lack of concern over social control by AI but more concern over discrimination, which was also significantly more common among all female students. Is it possible that many students simply reflected on their main personal worries in their everyday life rather than the ethical issues of AI systems? It is worth noting that all students who participated in the survey belonged to one of the two main colleges of the university: college of social sciences (195 students) and college of business management (272 students). The results could have been different if the study was done on students from engineering or computer science departments. Another limitation of this study was that the possible role of education in changing the knowledge and attitude of students was not measured through a before and after teaching experiment.

Declarations

Conflict of Interest The authors declare no competing interests.

References

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2019. Economic policy for artificial intelligence. *Innovation Policy and the Economy* 19: 139–159. https://doi.org/10.1086/699935.

Alaieri, Fahad, and André Vellino. 2016. Ethical decision making in robots: autonomy, trust and responsibility. In *Social Robotics. ICSR 2016. Lecture Notes in Computer Science, vol 9979*, eds. Arvin Agah, John-John Cabibihan, Ayanna M. Howard, Miguel A. Salichs, and Hongsheng He. Cham: Springer. https://doi.org/10.1007/978-3-319-47437-3_16.

Boddington, Paula. 2017. *Towards a code of ethics for artificial intelligence*. Cham: Springer. https://doi.org/10.1007/978-3-319-60648-4.

Bossman, Julia. 2016. Top 9 ethical issues in artificial intelligence. *World Economic Forum*. 21 October 2016. https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/. Accessed 31 Jul 2021.

Bartneck, Christoph, Christoph Lütge, Alan Wagner, and Sean Welsh. 2021. Privacy issues of AI. In An introduction to ethics in robotics and AI, 61–70. Cham: Springer. https://doi.org/10.1007/978-3-030-51110-4_8.

Curran, Dean, and Alan Smart. 2021. Data-driven governance, smart urbanism and risk-class inequalities: Security and social credit in China. *Urban Studies* 58(3): 487–506. https://doi.org/10.1177/0042098020927855.

Fiori, Marina, and Ashley K. Vesely-Maillefer. 2018. Emotional intelligence as an ability: theory, challenges, and new directions. In *Emotional Intelligence in Education: Integrating Research with Practice*, eds. Kateryna V. Keefer, James D.A. Parker, and Donald H. Saklofske, 23–47. Cham: Springer. https://doi.org/10.1007/978-3-319-90633-1_2.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, et al. 2018. AI4People, an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds & Machines* 28(4): 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Ghotbi, Nader, Manh Tung Ho, and Peter Mantello. 2021. Attitude of college students towards ethical issues of artificial intelligence in an international university in Japan. *AI & Society*. https://doi.org/10.1007/s00146-021-01168-2.

Hagerty, Alexa, and Igor Rubinov. 2019. Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. *arXiv*, 1907.07892v1. http://arxiv.org/abs/1907.07892. Accessed 31 Jul 2021.

Hara, Hiromi. 2018. The gender wage gap across the wage distribution in Japan: Within- and between-establishment effects. *Labour Economics* 53: 213–229. https://doi.org/10.1016/j.labeco.2018.04.007.

Hoffman, Samantha. 2018. Managing the state: social credit, surveillance and the CCP's plan for China. In *AI, China, Russia, and the global order: technological, political, global, and creative perspectives*, ed. Nicholas D. Wright, 42–47. Department of Defense, United States of America. http://nsiteam.com/social/wp-content/uploads/2019/01/AI-China-Russia-Global-WP_FINAL_forcopying_Edited-EDITED.pdf. Accessed 31 Jul 2021.

King, Thomas C., Nikita Aggarwal, Mariarosaria Taddeo, and Luciano Floridi. 2020. Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. *Science and Engineering Ethics* 26(1): 89–120. https://doi.org/10.1007/s11948-018-00081-0.

Korinek, Anton, Joseph E. Stiglitz. 2019. Artificial intelligence and its implications for income distribution and unemployment. In *The Economics of Artificial Intelligence: An Agenda*, eds. Ajay Agrawal, Joshua Gans, and Avi Goldfarb, 349–390. Chicago: University of Chicago Press.

Leslie, David. 2019. Understanding artificial intelligence ethics and safety: a guide for the responsible design and implementation of AI systems in the public sector. *SSRN* 3403301. https://doi.org/10.2139/ssrn.3403301.

Liang, Yuhua, and Seungcheol A. Lee. 2017. Fear of autonomous robots and artificial intelligence: Evidence from national representative data with probability sampling. *International Journal of Social Robotics* 9(3): 379–384. https://doi.org/10.1007/s12369-017-0401-3.

Lorenčík, D., M. Tarhaničová, and P. Sinčák. 2013. Influence of Sci-Fi films on artificial intelligence and vice-versa. In: *Proceedings of the IEEE 11th International Symposium on Applied Machine Intelligence and Informatics*, 27–31. https://doi.org/10.1109/SAMI.2013.6480990.

Mainichi. 2016. Insurance firm to replace human workers with AI system. *Mainichi*, 30 December 2016. https://mainichi.jp/english/articles/20161230/p2a/00m/0na/005000c. Accessed 31 Jul 2021.

Mikhaylenok, Oleg M., and Galina A. Malysheva. 2019. Political effects of social media in Russia [Политические эффекты социальных сетей в России]. Sociological Research [*Социологические исследования*] 2: 78–87. https://doi.org/10.31857/S013216250004012-6.

Ntoutsi, Eirini, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Esther Vidal, et al. 2020. Bias in data-driven artificial intelligence systems—an introductory survey. *WIREs Data Mining and Knowledge Discovery* 10(3): e1356. https://doi.org/10.1002/widm.1356.

O'Neill, C. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Penguin.

O'Sullivan, Shane, Nathalie Nevejans, Colin Allen, Andrew Blyth, Simon Leonard, Ugo Pagallo, et al. 2019. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *International Journal of Medical Robotics and Computer Assisted Surgery* 15(1): e1968. https://doi.org/10.1002/rcs.1968.

Ozawa, Harumi. 2021. Robot pets help ease virus isolation in Japan. *Japan Times*. 4 March 2021. https://www.japantimes.co.jp/news/2021/03/04/national/social-issues/robots-virus-isolation/. Accessed 31 Jul 2021.

Pantserev, Konstantin A. 2020. The malicious use of AI-based deepfake technology as the new threat to psychological security and political stability. In *Cyber defence in the age of AI, smart societies and augmented humanity*, eds. Hamid Jahankhani, Stefan Kendzierskyj, Nishan Chelvachandran, and Jaime Ibarra, 37–55. Cham: Springer. https://doi.org/10.1007/978-3-030-35746-7_3.

Peillex, Jonathan, Sabri Boubaker, and Breeda Comyns. 2019. Does it pay to invest in Japanese women? Evidence from the MSCI Japan empowering women index. *Journal of Business Ethics* 170(3): 595–613. https://doi.org/10.1007/s10551-019-04373-8.

Polyakova, Alina, and Chris Meserole. 2019. *Exporting digital authoritarianism: the Russian and Chinese models*. Washington DC: Brookings Institute. https://www.brookings.edu/research/exporting-digital-authoritarianism/. Accessed 31 July 2021.

Rodriguez-Bailon, Rosa, Boyka Bratanova, Guillermo B. Willis, Lucia Lopez-Rodriguez, Ashley Sturrock, and Steve Loughnan. 2017. Social class and ideologies of inequality: How they uphold unequal societies. *Journal of Social Issues* 73(1):99–116. https://doi.org/10.1111/josi.12206.

Schofield, Damian. 2018. Digital emotion: How audiences react to robots on screen. *Computer Applications* 5(1): 1–20. https://doi.org/10.5121/caij.2018.5101.

Stahl, Bernd C., Josephina Antoniou, Mark Ryan, Kevin Macnish, and Tilimbe Jiya. 2021. Organisational responses to the ethical issues of artificial intelligence. AI & Society. https://doi.org/10.1007/s00146-021-01148-6.

Tegmark, Max. 2017. Life 3.0: Being human in the age of artificial intelligence. New York: Knopf.

Wagner, Cosima. 2009. 'The Japanese way of robotics': interacting 'naturally' with robots as a national character? In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 510–515. https://doi.org/10.1109/roman.2009.5326221.

Walsh, Toby. 2018. Expert and non-expert opinion about technological unemployment. *International Journal of Automation and Computing* 15(5): 637–642. https://doi.org/10.1007/s11633-018-1127-x.

Wilkinson, Richard G., and Michael Marmot. 2003. *Social determinants of health: The solid facts*. Geneva: World Health Organization. https://www.euro.who.int/en/publications/abstracts/social-determinants-of-health.-the-solid-facts. Accessed 1 Aug 2021.

Winkle, Katie, and Paul Bremner. 2017. Investigating the real world impact of emotion portrayal through robot voice and motion. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 627–634. https://doi.org/10.1109/ROMAN.2017.8172368.

Yigitcanlar, Tan, Kevin C. Desouza, Luke Butler, and Farnoosh Roozkhosh. 2020. Contributions and risks of artificial intelligence (AI) in building smarter cities: insights from a systematic review of the literature. *Energies* 13(6): 1473. https://doi.org/10.3390/en13061473.

Zhang, Baobao, and Allan Dafoe. 2019. Artificial intelligence: American attitudes and trends. *SSRN* 3312874. https://doi.org/10.2139/ssrn.3312874.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.