



# Artificial intelligence ELSI score for science and technology: a comparison between Japan and the US

Tilman Hartwig<sup>1,2</sup> · Yuko Ikkatai<sup>2,4</sup> · Naohiro Takanashi<sup>3</sup> · Hiromi M. Yokoyama<sup>2</sup>

Received: 18 May 2021 / Accepted: 8 November 2021 / Published online: 22 January 2022  
© The Author(s) 2021

## Abstract

Artificial intelligence (AI) has become indispensable in our lives. The development of a quantitative scale for AI ethics is necessary for a better understanding of public attitudes toward AI research ethics and to advance the discussion on using AI within society. For this study, we developed an AI ethics scale based on AI-specific scenarios. We investigated public attitudes toward AI ethics in Japan and the US using online questionnaires. We designed a test set using four dilemma scenarios and questionnaire items based on a theoretical framework for ethics, legal, and social issues (ELSI). We found that *country* and *age* are the most informative sociodemographic categories for predicting attitudes for AI ethics. Our proposed scale, which consists of 13 questions, can be reduced to only three, covering ethics, tradition, and policies. This new AI ethics scale will help to quantify how AI research is accepted in society and which area of ELSI people are most concerned with.

**Keywords** Artificial Intelligence · Ethics · ELSI · Dilemma

## 1 Introduction

### 1.1 Development of artificial intelligence

We rely on artificial intelligence (AI) in academia and in our daily lives. The public debates over the ethics of AI in

media articles increased between 2013 and 2018, but only 11% of the articles mention ethical theories and principles based on an ethical framework. This study points out that a multifaceted approach is necessary to cover the social, ethical, and legal issues for AI technology (Ouchchy et al. 2020). The development of a tool for quantifying the ethics of AI is essential, especially for understanding the ethics-related attitudes of people toward AI research and to proceed with the discussions in society. However, such a tool has not been developed. In this study, we propose an AI ethics scale based on AI-specific scenarios.

### 1.2 Measuring ethics

Many studies have been conducted on how individuals learn to make moral judgments and how we think about ethics. Kohlberg's cognitive–developmental theory identified three developmental levels of moral reasoning: pre-conventional morality (level 1), conventional morality (level 2), and post-conventional morality (level 3). Moreover, each level has two substages. Level 1 includes obedience and punishment orientation (stage 1) and individualism and exchange (stage 2). Level 2 includes good interpersonal relationships (stage 3) and maintaining the social order (stage 4). Level 3 includes social contract and individual rights (stage 5) and universal

---

Tilman Hartwig and Yuko Ikkatai equal contribution.

✉ Hiromi M. Yokoyama  
hiromi.yokoyama@ipmu.jp

Tilman Hartwig  
hartwig@phys.s.u-tokyo.ac.jp

Yuko Ikkatai  
y.ikkatai@gmail.com

Naohiro Takanashi  
naohiro.takanashi@emp.u-tokyo.ac.jp

- <sup>1</sup> Institute for Physics of Intelligence, The University of Tokyo, Tokyo, Japan
- <sup>2</sup> Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), The University of Tokyo, Kashiwa, Japan
- <sup>3</sup> Executive Management Program Office, The University of Tokyo, Tokyo, Japan
- <sup>4</sup> Present Address: Faculty of Social Sciences, Kanazawa University, Kanazawa, Japan

ethical principles (stage 6). We moved up the stages in the developmental process (Kohlberg 1968; Crain 1985).

Other moral studies differentiate the ethics of autonomy, community, and divinity. The ethics of autonomy emphasizes the self; the ethics of community emphasizes the person as a member of a group; and the ethics of divinity focuses on people as spiritual or religious beings (e.g., Jensen 2011, n.d.; Shweder et al. 1997). Moral Foundation Theory identifies five elements that enable a discussion of morals related to ideologies (Haidt, 2012; Murayama and Miura 2019): harm/care (providing care without causing harm to others); fairness/reciprocity (maintaining fairness and reciprocity), ingroup/loyalty (keeping loyalty inside the group), authority/respect (respect for authority and hierarchy), and purity/sanctity (safeguarding purity and sanctity) (Haidt 2012; Haidt and Joseph 2004, 2007).

Morals and ethics of individuals have been measured both qualitatively (e.g., Moral Judgement Interview (MJI)) and quantitatively (e.g., Defining Issues Test (DIT)) (Elm and Weber 1994). Although there are some differences between the two techniques, a high correlation between MJI and DIT was reported (McGraw and Bloomfield 1987). One of the advantages of DIT is the independence of trained interviewers (Elm and Weber 1994). DIT measures the ability to make moral judgments, and was developed from Kohlberg's cognitive–developmental theory (Rest et al. 1974). For DIT, respondents read six scenarios with everyday ethical dilemmas. They answered 12 items corresponding to the six stages using a 5-point Likert scale and then ranked the four most important items. The proportion of items in postconventional morality (level 3) within the four items is calculated as an index p-score. DIT was later revised as DIT-2 with five different scenarios (Rest et al. 1999). DIT-related studies reported that the developmental stages with *age* were almost consistent with Kohlberg's theory in many countries (e.g., US, Rest et al. 1978; Japan, Yamagishi 1976). Yamagishi (1976) indicated a cultural difference: Japanese students move to stage 3 early and stay at stage 3 longer than US students. This result was discussed in association with a Japanese culture that emphasized “good interpersonal relationships” in stage 3.

DIT has been used in professional ethical education. For example, university students in the US took DIT before and after a course on engineering ethics. Their ability to make moral judgements significantly increased after the course (Self and Ellison 1998). Other fields, such as medicine, veterinary medicine, law, and dentistry, have also evaluated the effectiveness of their education curriculum (Bebeau 2002).

Field-specific ethics tests have been developed based on the idea that ethics should be measured within a practical context. The Professional Ethics in Pharmacy test (PEP test) was developed in Australia using a short-form DIT with three scenarios (Chaar 2009). The PEP test includes

three moral dilemma scenarios (over-the-counter products, morphine, and repeat prescriptions) that a pharmacist is likely to encounter. The respondents answered 12 items corresponding to three themes: business orientation, rules and regulations, and patients' rights. Kruijtbosch et al. (2019) conducted the PEP test in the Netherlands and reported a difference in the schematic structure from Australia. The Engineering and Science Issues Test (ESIT) was also derived from DIT-2, and includes six dilemma scenarios that a scientist or engineer is likely to encounter (Borenstein et al. 2010). Reidenbach and Robin (1988, 1990) developed an ethics scale in business. The respondents read three business moral dilemma scenarios (car, sales, and retail) and answered eight items using 7-point Thurstone scale for each one. This scale was developed based on five normative ethical philosophies: justice, relativism, deontology, egoism, and utilitarianism. Thirty-three items were reduced to eight items that included three dimensions: moral equity, relativism, and contractualism (Reidenbach and Robin 1990).

### 1.3 Theoretical framework to measure the ethics of AI

We devised a new scale to help us think about the ethics of AI within society. We used a framework of ethics, legal, and social issues (ELSI) from the field of society, technology, and science (STS) that emphasizes the relationship between science, technology, and society. ELSI was started in 1990 as an international research program (Ethical, Legal and Social Implications Research Program) of the Human Genome Project in the US. The associate director of the project, James Watson, declared that ELSI should be discussed for the project, and about 3% (later 5%) of the research budget was invested in ELSI research. The emphasis on ethics has been spreading to other fields.

We saw that the three dimensions of moral equity, relativism, and contractualism in business ethics (Reidenbach and Tobin 1990) had some overlap with the concept of ethics, legal, and social issues (ELSI) so, in this paper (Table 1), we proposed an ELSI scale based on business ethics. First, the dimension of *moral equity* in business ethics includes four items, which are the fundamental base for making decisions on moral issues. These items correspond to the ethical items in the ELSI scale. We replaced the business ethics items to match the context of ELSI. For example, based on the universal moral impact of AI, we replaced “acceptable/not acceptable to my family” with “positive/negative impact on the future.” Second, the dimension of *relativism* in business ethics includes two items that are inherited in social and cultural systems. These items correspond to the social items in the ELSI scale. In addition to these two items, we introduced two other items (“extremely favorable/unfavorable from a religious perspective” and “extremely favorable from

**Table 1** Correspondence between a business ethics scale and the ethics, legal, and social issues (ELSI) scale

Business ethics proposed by Reidenbach and Tobin (1990)		“Ethics, Legal and Social Issues” (ELSI) scale proposed by this work	
Moral equity	just/unjust fair/unfair morally/not morally right acceptable/not acceptable to my family	Ethical	makes/does not make sense at all very/not fair at all ethically very correct/incorrect positive/negative impact on the future
Relativism	culturally acceptable/unacceptable traditionally acceptable/unacceptable	Social	extremely favorable/unfavorable from a cultural perspective extremely favorable/unfavorable from a traditional perspective extremely favorable/unfavorable from a religious perspective extremely favorable/unfavorable from a public perspective
Contractualism	violates/does not an unspoken promise violates/does not violate an unwritten contract	Legal	explanation to society is sufficient/insufficient explanation to individuals is sufficient/insufficient transparency is sufficiently/insufficiently maintained policies and laws are sufficiently/insufficiently established

a public perspective”). The dimension of *contractualism* in business ethics includes two items about the social contract between business and society. These items correspond to the legal items in the ELSI scale. We listed four items focusing on accountability to society and its members.

We constructed four AI-specific scenarios based on recent literature:

The first scenario (scenario (a)) describes the use of AI to imitate deceased artists. AI scientists “revived” the performance of a famous singer, Misora Hibari, who died in 1989 and reproduced her voice with deep learning (Yamaha Corporation n.d.). She “appeared” on a stage as a hologram, performed one new song, and even gave a short message to the audience. The performance was broadcasted on a famous Japanese music TV program in 2019 and has triggered discussions in the Japanese media about the ethical limitations of AI research.

The second scenario (scenario (b)) describes the use of AI for product recommendations. Companies collect customer information and train AI recommendation systems to suggest additional buying options to customers. This practice increases profits for companies and can be convenient for customers. On the other hand, the customer’s personal information and metadata is used and sold (often without explicit consent) to manipulate customers’ behaviors and their preferences based on corporate intentions.

The third scenario (scenario (c)) describes the use of AI in autonomous weapon systems. With the help of AI, armies could theoretically reduce human casualties by avoiding unnecessary risks for soldiers. AI may even enable the military to attack with greater deadly force compared with conventional weapons. On the other hand, autonomous weapon systems can lower the threshold for killing, and liability issues in the case of a malfunction have not been resolved. Many AI and robotics researchers have signed an open letter that declared “starting a military AI arms race is a bad idea

and should be prevented by a ban on offensive autonomous weapons beyond meaningful human control” (Future of Life Institute 2015).

The fourth scenario (scenario (d)) describes the use of AI to prevent criminal activity. A database of someone’s behavior (such as GPS profile, credit card records, online search history), combined with security infrastructure (such as CCTV videos, police records), would allow AI to predict potential crimes. With further research, we could reduce crime rates by limiting the activities of people who might commit a crime. A survey reported that 70% of Japanese people favor predicting the signs of a potential crime using AI linked to surveillance camera images and information from witnesses who become suspicious (Hosotsubo et al. 2020). On the other hand, these surveillance techniques would fundamentally impact data protection and personal rights. Improving fairness and minimizing discrimination when using predictive algorithms is an ongoing field of research (Adebayo et al. 2015).

Public attitudes toward AI differ among countries. Funk et al. (2020) showed that the respondents who answered that AI has mostly been a good thing for society were 44% in the US and 65% in Japan. They pointed out that personal attributes, such as gender, age, education, income, as well as science and programming experience, also influenced this perception. In many countries, males more often than females, younger more often than older people, and educated people more often than the less well-educated tend to have positive attitudes toward AI. In one study in the US, male respondents with higher incomes and technology experience were found to be more supportive of high-level machine intelligence (Zhang and Dafoe 2019). A general interest in science and technology also influenced attitudes toward AI. Albarán et al. (2020) found that respondents in Spain showed negative attitudes toward AI if they were not interested in scientific discoveries and technological developments.

## 1.4 Research questions

Using these four AI scenarios, we investigated public attitudes towards the ethics of AI research based on three research questions:

RQ1: What differences can be found between Japan and the US?

RQ2: Which sociodemographic categories are most important to predict a person's attitude toward AI research?

RQ3: Which of the items regarding AI research ethics are most informative, and can we reduce this to fewer items without losing any information?

## 2 Methodology

We asked all the respondents to answer showing their level of agreement/disagreement with each scenario as well as to the 13 items using a 7-point Thurston scale. We also investigated how the 12 sociodemographic variables (10 personal attributes plus interest in science and technology, understanding of AI) influence their attitudes toward AI research ethics. This study was conducted in Japan and the US using online questionnaires.

### 2.1 Respondents

We conducted an online survey to investigate public responses to the ELSI score for AI. The authors contracted Cross Marketing Inc., a research company in Japan that collected the data from two populations using their data pool.

For the first population, people in Japan, the company sent an email to Japanese people who had previously registered. We collected data from 1,108 respondents (men = 556, women = 552) aged 20 to 69 years old (mean  $\pm$  SD = 44.8  $\pm$  13.8). These samples were prepared to match the current demographic profile of the Japanese population for age, gender, and location. The survey was conducted from September 2 to 4, 2020 in Japan.

For the second population, people in the US, the company sent an email to people who had previously registered. We collected data from 1,063 respondents (men = 533, women = 530) aged 20 to 69 years old (mean  $\pm$  SD = 44.4  $\pm$  14.3). These samples were prepared to match the current demographic profile of the US's population for age, gender, and location. The survey was conducted from September 7 to 15, 2020 in the US. This study received approval from the Institutional Ethics Committee of the University of Tokyo (No. 20–120).

### 2.2 Procedure

Questionnaire items. The survey consisted of demographic variables (1, 2, 3) and the questionnaire items for each scenario (4).

- (1) Age, gender, location, marital status, occupation, house income, number of children, education, political party affiliation, and hours of PC/smartphone use (see Appendix 1).
- (2) Level of interest in science and technology (S&T). VSEG, a segmentation method was used to classify the level of interest in science and technology. VSEG consists of three items (ST\_Q1–Q3). The respondents' attitudes are classified into three groups (with interest, with potential interest, with low interest) by combining the responses to the three items (Victorian Department of Innovation, Industry and Regional Development 2011; Goto et al. 2014, Table 2).
- (3) ST\_Q1. How much are you interested in science and technology? (1: Very interested, 2: Quite interested; 3: Neither interested not disinterested; 4: Not very interested; 5: Not interested at all, 6: Don't know)
- (4) ST\_Q2. Do you actively search for information about science and technology? (1: Yes; 2: No, 3: Don't know)
- (5) ST\_Q3: When you have looked for information about science and technology in the past, have you generally been able to find what you were looking for? (1: Yes, and it tends to be easy to understand; 2: Yes, but it is often difficult to understand; 3: No, I often can't find what I am looking for; 4: Don't know)
- (6) Understanding of AI. We prepared three questions to measure the level of AI knowledge. These questions were reviewed by AI experts and graduate students studying AI in Japan.
- (7) Quiz\_Q1. Which of the following options is the most appropriate explanation of AI as of today? (1: A robot that thinks and acts on its own, without human assistance; 2: A program that makes decisions based on

**Table 2** Three groups in the Victorian Segment (VSEG)

	Group with interest	Group with potential interest	Group with low interest
ST_Q1	1 or 2	1, 2, 3, 4, or 5	4 or 5
ST_Q2	1	1 or 2	2
ST_Q3	1, 2 or 3	–	–

Respondents who did not meet the above criteria were classified as N/A. A survey of the Japanese public in 2013 showed that the percentages of groups were 16.1% with interest, 61.4% with potential interest, and 22.6% with low interest (PESTI n.d.)

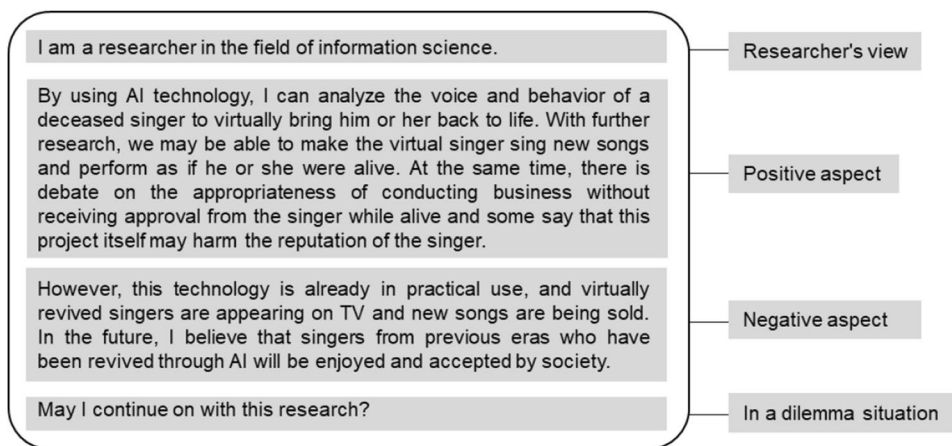
- learning results; 3: A computer that interacts with people; 4: A new type of smartphone).
- (8) Quiz\_Q2. Which of the following options is the most appropriate explanation of what AI can do as of today? (1: It makes moral decisions on its own; 2: It understands and interprets human languages; 3: It develops software on its own; 4: It has free will).
  - (9) Quiz\_Q3. Which of the following options is the most appropriate explanation of AI developers as of today? (1: The government is developing AI; 2: Information scientists and researchers are developing AI; 3: Computer programs are developing AI without human intervention; 4: Everyone is developing AI using smartphones).
  - (10) Items for the scenarios. We prepared four scenarios ((a)–(d)), each consisting of a short paragraph (see Appendix 2). The scenarios described the use of AI for AI-generated singers (scenario (a)), AI customer

purchases (scenario (b)), AI autonomous weapons (scenario (c)), and AI prediction of criminal activities (scenario (d)). The scenarios consisted of a description of a researcher and included beneficial and anxiety-inducing aspects of AI. The researcher is facing an ethical dilemma: whether or not to continue with research (Fig. 1).

We asked the respondents to answer the following 13 items (Q0–Q12) after reading each scenario. We prepared the scenarios both in Japanese and English using double-back translation from Japanese to English and English to Japanese to ensure accuracy.

- Q0. “Do you agree or disagree with this research?” was asked. The responses were rated on a seven-point scale from “I strongly agree (= 1)” to “I strongly disagree (= 7).

**Fig. 1** Scenario (a) describes beneficial and anxiety-inducing factors as well as an ethical dilemma



**Table 3** Questionnaire design of Q1–Q12

	1	2	3	4	5	6	7
Q1. Very fair							Not fair at all
Q2. Makes a lot of sense							Does not make sense at all
Q3. Ethically very correct							Ethically very incorrect
Q4. Positive impact to the future							Negative impact to the future
Q5. Extremely favorable from a cultural perspective							Extremely unfavorable from a cultural perspective
Q6. Extremely favorable from a traditional perspective							Extremely unfavorable from a traditional perspective
Q7. Extremely favorable from a religious perspective							Extremely unfavorable from a religious perspective
Q8. Extremely favorable from a public perspective							Extremely unfavorable from a public perspective
Q9. Explanation to society is sufficient							Explanation to society is insufficient
Q10. Explanation to individuals is sufficient							Explanation to individuals is insufficient
Q11. Transparency is sufficiently maintained							Transparency is insufficiently maintained
Q12. Policies and laws are sufficiently established							Policies and laws are insufficiently established

The respondents chose the most appropriate option from among seven options

- Q1–Q12. “What do you think of this research?” was asked for 12 items. The responses were rated on a seven-point scale. Only the sides of the scales were labelled: Table 3).

### 2.3 Analysis

We first calculated the item-total correlation and the item-remainder correlation to ensure that all 12 questions contribute to measuring a consistent scale and to test if the questionnaire can be reduced. Then, we used decision trees and calculated feature importance to find the most important sociodemographic categories (for RQ2) as well as to reduce the catalogue (list) of items (for RQ3). Decision tree analysis is a logical graph structure based on labelled training data that optimally differentiates between labelled classes (Quinlan 1986). Decision trees are constructed from the top down, i.e., from a head node down to leaf nodes, which contain only one class of data, in the ideal case.

Decision trees are most stable on binary classification tasks; therefore, we prepared the data in the following way. For each of the four scenarios and each of the 13 items (Q0–Q12), we divided the seven answer options (option 1 to 7) into two groups: positive answer (option 1 and 2) and negative answer (option 6 and 7). Ignoring answer options 3 through 5 reduced the sample size by 60% and made the final prediction more robust, i.e., provides a higher validation

accuracy on a blind test set. We have verified that the final results are not affected by this data preprocessing technique.

For example, we took Q3 (Ethically very correct/incorrect) of scenario (b) and ignored all responses that answered option 3 through 5 for this item. Taking the two extreme answers (options 1, 2 and options 6, 7) made the decision tree more robust. Then, we divided these selected answers into 10% test data and 90% training data to train the decision tree. The goal was to predict “positive” or “negative” attitudes toward AI research based on the sociodemographic data (for RQ2) and based on the answers to the other items and scenarios (for RQ3).

Construction of a decision tree is deterministic: the head node contains both classes. The goal for us was to split the data based on the available information to obtain purer nodes, where one class dominates. The only rule for building a decision tree is that every new split should minimize the information entropy with respect to the assigned classes (positive/ negative). An example of a decision tree is given in Fig. 2. Altogether, we trained 52 (13 items × 4 scenarios) independent decision trees for sociodemographic data (for RQ2) as well as the remaining AI items (for RQ3). The prediction accuracy of the decision trees is in the range of 70–90%, for most models.

One primary outcome of a decision tree is the feature importance. The feature importance of one item is obtained by summing the entropy decrease over all nodes, where this information is used, weighted by the fraction of answers that are split in this node. The higher the value of the feature

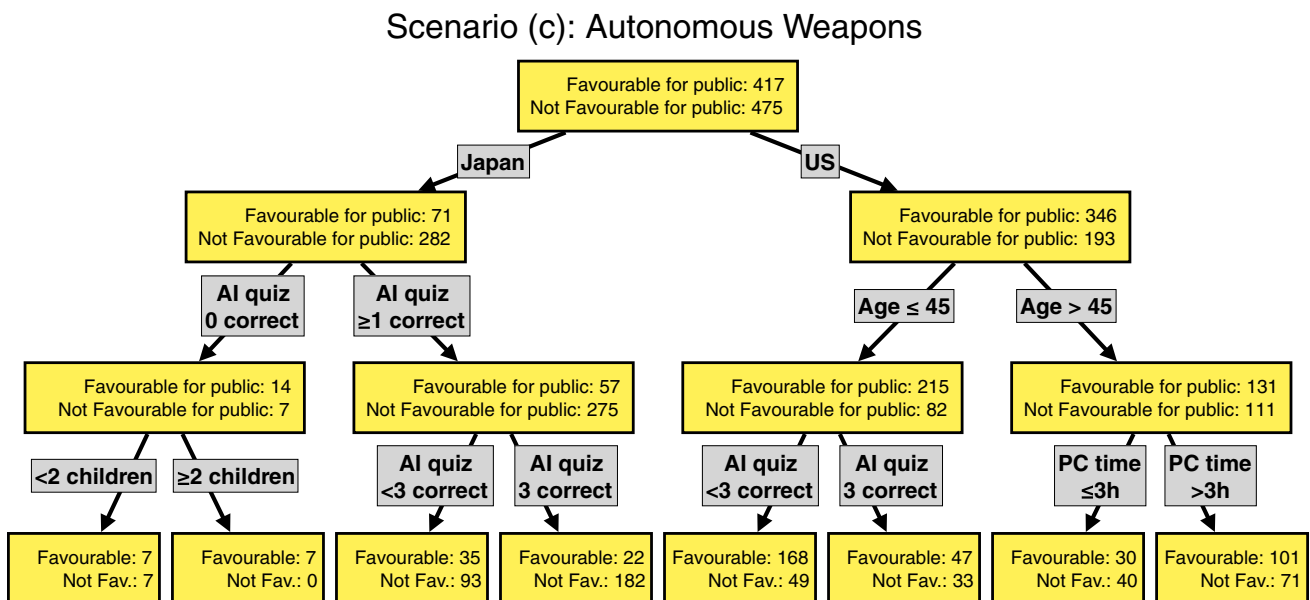


Fig. 2 Example decision tree for Q8 (extremely favorable/unfavorable from a public perspective) in scenario (c), for which we use sociodemographic data to classify the answers. This optimal decision tree achieves a prediction accuracy of 85% on the blind test set

importance, the more it contributes to the classification of the target.

For example, the top node in Fig. 2 has an information entropy of 0.997. In the first layer, the left and right nodes have the information entropies 0.724 (Japan) and 0.941 (US). Their combined, weighted information entropy is, therefore, 0.855. Therefore, in this split the socioeconomic item *country* would gain a feature importance of  $0.997 - 0.855 = 0.142$ .

The feature importance quantifies how informative an item is for the decision process (i.e., predicting whether a person would reply positively or negatively to a particular item for a specific scenario). We weighted the feature importance from all 52 decision trees with the accuracy of each decision tree. This modification gives weight to more robust and reliable decision trees. For the analysis, we have used Scikit-learn: *Scikit-learn: Machine Learning in Python* (Pedregosa et al. 2011).

### 3 Results

#### 3.1 RQ1: differences between Japan and US

First, we show the level of interest in science and technology (S&T) and the understanding of AI, since these items indicate the most prominent difference. While people in the US seem to have a higher interest in S&T, people in Japan scored higher on our AI quiz (Fig. 3). On average, US respondents had, among the three quiz items,  $1.9 \pm 0.9$  answers correct. On the other hand, Japanese survey respondents answered  $2.2 \pm 1.0$  items correctly. The fraction of respondents who answered all three AI quiz items correctly is almost twice as high in Japan (545/1108) compared with the US (301/1063).

The second remarkable difference is the use of PCs or smartphones. People in the US use their smartphones or PCs

on average  $4.9 \text{ h} \pm 3.3 \text{ h}$  daily; however, people in Japan use them for  $3.1 \text{ h} \pm 2.8 \text{ h}$ .

Third, we show the mean values from Japan (JP) and the US (US) and their differences (JP-US) in Table 4. A positive value for the difference (JP-US) indicates that people in Japan are on average more negative toward AI research in this category. The last column (average difference) shows the difference between Japan and the US, averaged over the four scenarios. For 50 out of 52 items, people in Japan answered on average with higher score values compared with their US counterparts. This indicates that people in Japan, in the four tested scenarios, seem to have a more negative attitude toward AI research. Interestingly, there were only two cases, where people in the US were on average more negative toward AI research: Q0 (“Do you agree or disagree with this research?”) in the AI-singer scenario (a) and the AI-prevention of criminal activities scenario (d). In addition, there is a general trend that people in Japan prefer to give the central (neutral) answer. From all 52 AI ethics items (13 items  $\times$  4 scenarios), the Japanese respondents answered “option 4” 36% of the time. On the other hand, US respondents answered “option 4” in only 20% of the cases. While we found different trends between the countries, there were no statistically significant differences ( $> 1$  sigma) between Japan and the US in the responses to all 13 items for the scenarios, as measured by the mean and standard deviation of the answers.

#### 3.2 RQ2: most important sociodemographic categories

We analyzed the feature importance of the entire sample. Specifically, we tried to predict whether the respondents would reply positively (option 1 and 2) or negatively (option 6 and 7) to the AI ethics items based on their 11 items of

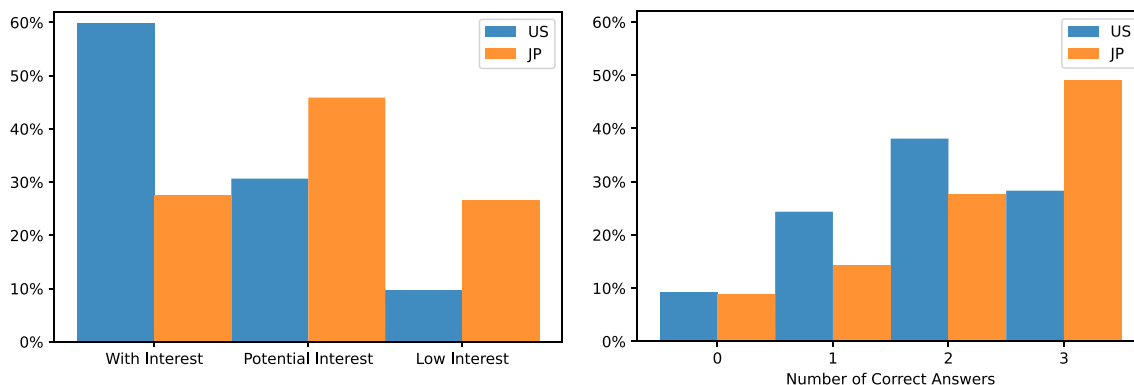


Fig. 3 Percentages of level of interest in science and technology (left) and correct answers on the AI quiz (right). Orange bars show Japan (JP) and blue bars show the US

**Table 4** Mean and differences of the responses to each scenario item

	Scenario (a)				Scenario (b)				Scenario (c)				Scenario (d)				Average difference		
	JP		US		JP-US		US		JP		US		JP-US		JP			US	
	JP	US	JP-US	US	JP	US	JP-US	US	JP	US	JP-US	US	JP	US	JP-US	JP		US	JP-US
Q0	4.08	4.17	-0.09	3.68	3.85	0.17	4.84	3.92	4.1	4.31	0.92	4.1	4.31	0.20	4.1	4.31	0.21	-0.21	0.20
Q1	4.01	3.78	0.23	3.23	3.87	0.64	4.53	3.43	4.04	3.83	1.1	4.04	3.83	0.55	4.04	3.83	0.21	0.21	0.55
Q2	4.07	3.68	0.39	2.97	3.71	0.74	4.61	3.13	3.94	3.49	<b>1.48</b>	3.94	3.49	0.77	3.94	3.49	0.45	0.45	0.77
Q3	4.25	4.18	0.07	3.53	3.97	0.44	4.81	3.73	4.16	4.06	1.08	4.16	4.06	0.42	4.16	4.06	0.1	0.1	0.42
Q4	3.9	3.78	0.12	3.3	3.71	0.41	4.76	3.57	3.87	3.7	1.19	3.87	3.7	0.47	3.87	3.7	0.17	0.17	0.47
Q5	3.98	3.79	0.19	3.48	3.87	0.39	4.71	3.68	4.08	3.99	1.03	4.08	3.99	0.43	4.08	3.99	0.09	0.09	0.43
Q6	4.16	4.08	0.08	3.65	4.02	0.37	4.72	3.8	4.19	4.12	0.92	4.19	4.12	0.36	4.19	4.12	0.07	0.07	0.36
Q7	4.29	4.17	0.12	3.77	4.07	0.3	4.75	4.04	4.24	4.11	0.71	4.24	4.11	0.32	4.24	4.11	0.13	0.13	0.32
Q8	4.05	3.69	0.36	3.42	3.91	0.49	4.72	3.6	3.98	3.92	1.12	3.98	3.92	0.51	3.98	3.92	0.06	0.06	0.51
Q9	4.29	3.75	0.54	3.45	4.2	0.75	5.01	3.63	4.49	3.89	1.38	4.49	3.89	0.82	4.49	3.89	0.6	0.6	0.82
Q10	4.35	3.75	<b>0.60</b>	3.45	4.34	<b>0.89</b>	4.99	3.59	4.55	3.89	1.4	4.55	3.89	<b>0.89</b>	4.55	3.89	<b>0.66</b>	<b>0.66</b>	<b>0.89</b>
Q11	4.27	3.8	0.47	3.58	4.37	0.79	4.97	3.75	4.56	4.02	1.22	4.56	4.02	0.76	4.56	4.02	0.54	0.54	0.76
Q12	4.46	4.01	0.45	3.68	4.43	0.75	5.09	3.84	4.68	4.09	1.25	4.68	4.09	0.76	4.68	4.09	0.59	0.59	0.76

Scenario (a) is the AI-generated singer, scenario (b) is AI-customer recommendations, scenario (c) is AI-autonomous weapons, and scenario (d) is AI-prediction of criminal activities. Boldface highlights the maximum difference for each scenario

sociodemographic data. We then renormalized the sum of the feature importance to 1 and list them in decreasing order: country 30.7%, age 29.5%, number of correct items in AI quiz (understanding of AI) 18.0%, occupation 6.5%, number of children 4.9%, household income 4.3%, PC/smartphone use time 2.1%, level of interest in science and technology (S&T) 1.8%, education 1.1%, gender 0.7%, and marital status 0.3%. This demonstrates that *country* was the most crucial item for predicting opinions of AI ethics. Age and respondents' understanding of AI were also important indicators. Gender and marital status were not essential indicators. The level of interest in S&T and education also do not play an important role.

### 3.3 RQ3: reduced catalogue of items

We first calculated the item-total correlation and the item-remainder correlation to ensure the robustness of the proposed scale and to check if we can reduce the number of questions. The Pearson's correlations between the total score of 12 items (from 12 to 84) and each item (from 1 to 7) was calculated per scenarios. The coefficients ranged from 0.77 to 0.92, suggesting that each of the questions are highly correlated to the total score (Appendix 3). It demonstrates that the quantities measured by Q1–Q12 and the sum of the 12 items are consistent.

In addition, we calculated item-remainder correlation. The Pearson's correlations between the total score of each item across the four scenarios (from 4 to 28) and each item in each scenario was calculated. Due to the small number of items (four scenario items), we chose item-remainder correlation to avoid the same item being included in the calculation of correlations. The coefficient ranged from 0.58 to 0.74 (Appendix 3). This test confirms the robustness of our scale across scenarios. However, all correlation coefficients are > 0.5, no question has significantly smaller correlations, and no coefficient is < 0.3. While this shows that our scale is consistent and robust, this correlation analysis alone does not identify questions that could be removed in order to reduce the size of the questionnaire.

Therefore, we used decision trees to predict the answers to one AI ethics item (13 items × 4 scenarios) based on the answers to the remaining 51 items. Figure 4 shows the feature importance of the 13 items to predict attitudes toward AI ethics. The value of the feature importance is largest in Q3 (Ethically very correct/incorrect), Q6 (Extremely favorable/unfavorable from a traditional perspective), and Q12 (Policies and laws are sufficiently/insufficiently established), which suggests that these three items are most important to predict opinions about AI ethics. While Q1 (Very fair/not fair at all) also provides a high feature importance, we did not include it here. We will demonstrate in the next paragraph that the inclusion of Q1 is not necessary, provided that



Q3 is selected. Therefore, these three items (Q3, Q6, Q12) contain the most information, and the answers to other items can be predicted based on the answers to these three items.

To independently confirm this set of reduced three items (Q3, Q6, Q12), we also applied linear regression to our data. We first averaged the answers to the 13 items over the four scenarios, then we performed a linear regression to predict the answers to the remaining 10 items, based on three items (Qi, Qj, Qk). We have tried all combinations of three items and noted the resulting R values of the fit accuracy. We then selected the top 10% (best) models and analyzed which triplets of items these models were using. The items used for the top 10% of models were not randomly distributed but instead contained one item from Q0–Q4, one item from Q5–Q8, and one item from Q9–Q12 (Fig. 5). This shows that the best models to predict the remaining 10 items, based on just three items, tended to use one item from each of the three categories (Q0–Q4, Q5–Q8, Q9–Q12). This confirms our previous decision tree-based finding that items Q3, Q6, and Q12 can be considered the most informative items to reduce the scale to just three items. These three most informative items naturally cover the three main blocks of items, which could be divided into ethics (Q0–Q4), society (Q5–Q8), and legal (Q9–Q12).

### 4 Findings

RQ1: People in the US have a higher interest in science and technology (S&T) but perform less well in the AI quiz than people in Japan. Our data shows that people in Japan, on average, hold more negative views toward AI research in the four tested scenarios.

RQ2: Based on our data, country is the most important socioeconomic indicator to predict attitudes toward AI ethics (30.7%). The next important categories are age (29.5%) and the number of correct answers in the AI quiz (18.0%).

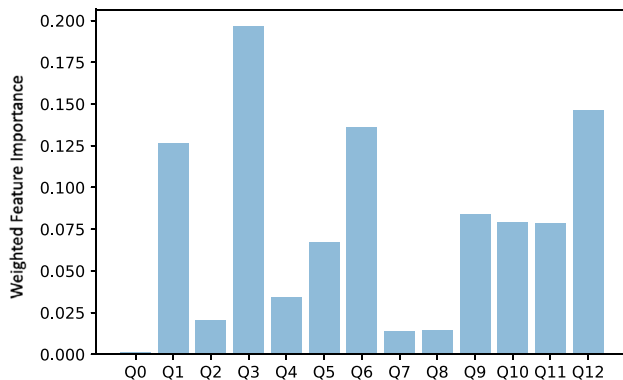


Fig. 4 Feature importance of the 13 items to predict the attitudes toward AI ethics

The remaining sociodemographic information has only a small feature importance:  $\leq 7\%$ .

RQ3: We find that the items Q3, Q6, and Q12 provide the highest feature importance and can be considered the three items with the most significant information gain, compared with the remaining items. We confirm this result with an independent regression model, which also shows that it is essential to have one item of each of the three groups: ethics (Q0–Q4), society (Q5–Q8), and legal Q9–Q12.

### 5 Discussion

In this study, we quantitatively investigated public attitudes toward AI research ethics using a set of dilemma scenarios. We found that items Q3, Q6, and Q12 provide the highest feature importance and can, therefore, be considered the three items with the largest information gain, compared with the other items. We confirmed this result with an independent regression model, which also shows that it is important to have one item of each of the three groups: ethics (Q0–Q4), society (Q5–Q8), and legal Q9–Q12. Using Principal Component Analysis (PCA) we confirmed that the dimensionality of our data is reducible, and only three dimensions can recover over 80% of the variance. The three most informative items that we suggest as the basis for a reduced questionnaire are Q3 (Ethically very correct/incorrect), Q6 (Extremely favorable/unfavorable from a traditional perspective), and Q12 (Policies and laws are

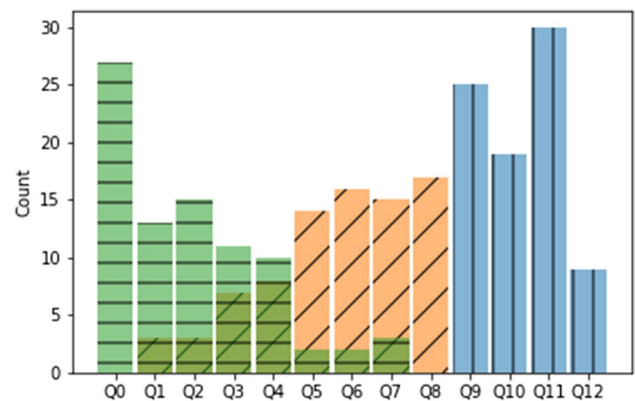


Fig. 5 Linear regression result: We tested all combinations for three out of 13 questions to see how well each triplet could predict the remaining 10 questions. This histogram shows how often each of the 13 questions provides an R value in the top 10% of all possible combinations. The colors indicate the smallest item for each triplet (green, horizontal hatch), the middle item (orange, diagonal hatch), and the highest item (blue, vertical hatch). For example, the triplet Q3, Q6, Q12 provides an excellent R value. Then, Q3, as the smallest of these numbers, is accounted for by the green histogram, Q6 for the orange histogram, and Q12 for the blue histogram

sufficiently/insufficiently established). The response to these three items summarizes the respondents' attitudes toward the ELSI (ethics, legal, and social issues) of AI. This analysis demonstrates that asking these three questions can identify which area of ethics, social, and legal issues people are most concerned with. This short version of the ELSI score, a set of dilemma scenarios and only three items, would be an easy questionnaire for many people to answer. The short version would make the results more transparent and easier to interpret.

In our analysis, we found that *country* (feature importance: 30.7%) was the most important category to predict attitudes toward AI ethics. This means that knowing a person's home country may be more important than knowing that person's age, education, or profession. There are several reasons why the answers between Japan and the US differ, such as items that depend on the country-specific laws and scientific culture (Q9–Q12).

The second category is *age* (29.5%), a possible reflection of older people being less exposed to AI as they grew up and are, therefore, more skeptical than younger people who use AI every day. A previous study reported that younger adults showed positive attitudes toward the use of AI in society (Funk et al. 2020).

The third category is the *AI quiz* (18.0%). People who knew more about AI could be more skeptical about it. The benefits of AI technology are evident to most people; however, the negative aspects of AI technology are often more subtle and require a deeper understanding of biases in data, human-constructed loss functions to train AI, or the value of personal data. The more we understand about science and technology (S&T), the more cautious we could possibly become toward it.

Among the least important categories are *education* (1.1%) and *gender* (0.7%). However, another study reported that educated people, and men more than women, are likely to agree that the development of AI is a good thing for society (Funk et al. 2020; Zhang and Dafoe 2020). The feature importance that we used for this ranking does not conserve correlations between categories. Therefore, it is possible that the influence that education has on AI ethics is already included in the AI quiz variable, which has a higher feature importance.

We found that interest in S&T was higher in the US than in Japan. This tendency was also found in previous studies. For example, a report from 2001 showed that an index of interest in S&T was higher in the US (67 points) than in Japan (50 points) (Ministry of Education, Culture, Sports, Science and Technology 2002). In the World Values Survey Wave 7 (2017–2020), the respondents who agreed (those who chose option 6–10 (= a lot better off) among 10 scales) that the world is better off because of S&T was higher in the US (75%) than Japan (69%) (World Values Survey

Association n.d.). People in the US have a higher interest in S&T but perform less well in the AI quiz than people in Japan. This shows that increased knowledge of AI does not necessarily mean an increased interest in S&T (Japan), and, vice versa, a high interest in S&T does not necessarily lead to knowledge of AI (US).

Another notable difference between Japan and the US was found in the responses to the dilemma scenarios. On average, people in Japan have a more negative attitude toward AI research in the four tested scenarios. One possible explanation for this difference is the mean age of the citizens: Japan's ageing population (*age* was the second most important category) may be less receptive to new technologies. Among the respondents aged 50 or older, daily PC use time was  $4.0 \text{ h} \pm 3.1 \text{ h}$  in the US and only  $2.6 \text{ h} \pm 2.5 \text{ h}$  in Japan. In both countries, these figures are lower than the average number for all age groups:  $4.9 \text{ h} \pm 3.3 \text{ h}$  in the US and  $3.1 \text{ h} \pm 2.8 \text{ h}$  in Japan.

The most remarkable difference appears for Q1 in scenario (c): AI autonomous weapons, which is more strongly supported in the US, and US respondents answered on average more than 0.92 points higher (in favor) on the scale from 1 to 7 (Table 4), compared with their Japanese counterparts. These attitudes may be influenced by the pacifist constitution of Japan ("The Japanese people forever renounce war as a sovereign right of the nation and the threat or use of force as a means of settling international disputes"), and on the fact that the US has intervened in various countries around the world. In other words, autonomous weapon systems could bring a more direct benefit for US society than Japanese society. The fact that Japanese AI guidelines do not focus on using AI-based autonomous weapons (Ema 2017) could be a reflection of this situation.

The items for which we found the most prominent differences between Japan and the US (averaged over the scenarios) are Q9 to Q12 (legal items). These items are intrinsically country-specific and can be influenced by the country's support and infrastructure for science outreach programs (Q9–Q11) or directly by regulations and how these are communicated to the public (Q12).

Item Q3 had to be included in the reduced questionnaire, because AI ethics is the overall narrative of this scale and the scenarios are constructed to show ethical dilemmas. Q6 (tradition) measures how open people are toward new technologies, which plays a significant role in their attitudes toward AI ethics. The importance of Q12 shows that people see a connection between new technologies and regulations: either that new technologies should be explored freely for the benefit of science or that politicians and other public stakeholders should regulate new technologies to mitigate potential harm. In either case, category Q12 can capture these sentiments that would

otherwise not be measured effectively by any other item. In a future study, we will test the hypothesis that this reduced set of items provides similar results to our complete set of 13 items.

This demonstrates that the final scale must be based on several (more than one) scenarios. For example, to predict what the respondents answered for scenario (a), the most informative were usually from the remaining 12 questions on scenario (a) and only a few times from the remaining 39 questions (3 scenarios  $\times$  13). There is no average or general attitude toward AI research ethics that can be generalized from one scenario to all other.

## 5.1 Perspectives

Various AI guidelines have been published (e.g., Hagen-dorff 2020; Fjeld et al. 2020). For example, Fjeld et al. (2020) reviewed 36 guidelines from various countries. They extracted eight key themes: privacy, accountability, safety and security, transparency and explainability, fairness and nondiscrimination, human control of technology, professional responsibility, and promotion of human values. These eight themes represent the global issues surrounding AI ethics. Our ELSI scale does not assess the public attitudes for general AI ethics, rather it visualizes the problems based on more concrete scenarios. In Europe and the US, the concept of RRI (research responsible innovations) based on ELSI or ELSA (ethics, legal, social aspects) is widely used for the development of science and technology (von Schomb-erg 2011). RRI emphasizes discussions that involve various stakeholders in society. To start a debate on the ethics of AI in society, it is essential to visualize the public attitudes in each country. Our ELSI scale allows us to initiate a dialogue with the public about AI-related developments that may occur in the future. As AI is now progressing rapidly, it is urgent to start this discussion in society. However, we further need to investigate this ELSI scale for reliability and validity. Therefore, this study is the first step to quantify public attitudes for the ethics of AI to enable future international comparisons.

## 5.2 Limitations and caveats

While decision trees are more transparent and inspectable than other machine learning techniques, their construction based on information entropy is not very intuitive and dividing the data randomly into a training and blind test set makes the outcome nondeterministic. We have tested that the main results are robust with respect to the stochastic nature of this approach.

We have analyzed four scenarios for AI research. However, our conclusions cannot be seen as general trends for AI research, rather the results should be interpreted as specific to these four scenarios. Increasing the number of scenarios in the future would mitigate this limitation.

We used a quantitative questionnaire study; the respondents could only answer the questionnaire items that we prepared based on a discrete scale. Such an approach has limitations and there may be nuances that we could not detect using a quantitative approach. A future qualitative study would further support our discussion.

We found that knowledge of AI is an essential factor that influences a person's attitude toward AI research. We measured the AI knowledge of respondents based on three questions with four multiple choice answers each in an AI quiz. The phrasing of these items and assessing a person's AI knowledge is difficult in such a dynamically evolving field as AI. Since we found that the AI quiz was important, more time needs to be invested in designing a general scale to evaluate a person's level of AI knowledge.

The answers for this study were collected via an online survey, and the oldest respondents were 69 years old. Since we found that *age* is an essential indicator of a person's attitude toward AI research, it is important to also include respondents who are older than 70 years in future surveys.

## 6 Conclusion

We investigated public attitudes for AI ethics in Japan and the US based on a theoretical framework of ethics, legal and social issues (ELSI). We found that *country* is the most important category to predict someone's attitudes toward AI ethics. We also found that the three items (Q3: ethical, Q6: traditional, Q12: policies and laws) are most informative, and they cover ethics (Q0–Q4), society (Q5–Q8), and legal (Q9–Q12). This set of dilemma scenarios and questionnaire items would be helpful, for instance, for AI researchers to analyze how their research is accepted in society and which area (ethics, legal, and society) people have concerns with. In future studies, we will apply this approach to other new technology and other countries to investigate the validity and limitations.

## Appendix 1

See appendix Table 5.

**Table 5** Responses to demographic variables

	US		Japan	
Gender	Men	533	Men	556
	Women	530	Women	552
Living area/states	WA	32	Hokkaido	136
	OR	21	Aomori	16
	CA	128	Iwate	21
	AK	5	Miyagi	39
	HI	6	Akita	8
	MT	3	Yamagata	18
	ID	6	Fukushima	25
	WY	0	Ibaraki	5
	NV	12	Tochigi	4
	UT	12	Gunma	4
	CO	21	Saitama	15
	AZ	29	Chiba	22
	NM	4	Tokyo	66
	ND	1	Kanagawa	27
	SD	2	Niigata	14
	NE	4	Toyama	6
	KS	7	Ishikawa	5
	MN	16	Fukui	2
	IA	10	Yamanashi	6
	MO	22	Nagano	13
	WI	18	Gifu	12
	IL	45	Shizuoka	21
	MI	38	Aichi	69
	IN	26	Mie	9
	OH	66	Shiga	6
	OK	7	Kyoto	19
	TX	45	Osaka	69
	AR	12	Hyogo	24
	LA	6	Nara	6
	KY	7	Wakayama	6
	TN	16	Tottori	14
	MS	8	Shimane	8
	AL	14	Okayama	38
	WV	1	Hiroshima	55
VA	17	Yamaguchi	21	
NC	19	Tokushima	23	
SC	16	Kagawa	36	
GA	25	Ehime	55	
FL	54	Kochi	22	
DE	4	Fukuoka	64	
MD	9	Saga	7	
DC	3	Nagasaki	17	
NY	104	Kumamoto	18	
PA	62	Oita	9	
NJ	37	Miyagi	7	
NH	7	Kagoshima	12	
VT	2	Okinawa	9	

**Table 5** (continued)

	US	Japan		
	ME	8	–	–
	MA	27	–	–
	RI	5	–	–
	CT	14	–	–
Marriage	Single/unmarried	506	Single/unmarried	587
	Married/civil partnership	557	Married/civil partnership	521
Occupation	Company employee	309	Company employee	364
	Company officer or manager	91	Company officer or manager	30
	Civil servant or Association staff member	24	Civil servant or Association staff member	69
	Self-employed	96	Self-employed	59
	Freelancer or professional	18	Freelancer or professional	39
	Dispatch or contract employee	8	Dispatch or contract employee	63
	Part-time worker	67	Part-time worker	157
	Elementary school student and below	3	Elementary school student and below	0
	Junior high school student	1	Junior high school student	0
	High school student	2	High school student	0
	Preparatory school student	0	Preparatory school student	0
	Vocational school student, junior college student, university student, or graduate student	21	Vocational school student, junior college student, university student, or graduate student	35
	Housewife or househusband	86	Housewife or househusband	164
	Unemployed	235	Unemployed	114
	Other	102	Other	14
House income	< 10,000USD	70	1,000,000JPY	45
	10,000USD–20,000USD	104	1,000,000JPY–2,000,000JPY	56
	20,000USD–30,000USD	104	2,000,000JPY–3,000,000JPY	115
	30,000USD–40,000USD	105	3,000,000JPY–4,000,000JPY	114
	40,000USD–50,000USD	97	4,000,000JPY–5,000,000JPY	131
	50,000USD–60,000USD	107	5,000,000JPY–6,000,000JPY	113
	60,000USD–70,000USD	79	6,000,000JPY–7,000,000JPY	78
	70,000USD–80,000USD	65	7,000,000JPY–8,000,000JPY	52
	80,000USD–90,000USD	46	8,000,000JPY–9,000,000JPY	52
	90,000USD–100,000USD	44	9,000,000JPY–10,000,000JPY	43
	100,000USD–120,000USD	57	10,000,000JPY–12,000,000JPY	35
	120,000USD–150,000USD	59	12,000,000JPY–15,000,000JPY	24
	150,000USD–200,000USD	51	15,000,000JPY–20,000,000JPY	15
	200,000USD >	41	20,000,000JPY >	10
	Do not know	13	Do not know	113
	Do not want to answer	21	Do not want to answer	
Number of children	0	594	0	665
	1	240	1	186
	2	155	2	161
	3	48	3	61
	4 or more	26	4 or more	35

Table 5 (continued)

	US	Japan		
Education	Elementary school / Junior high school	14	Elementary school / Junior high school	30
	High school	312	High school	315
	Junior college / Vocational school	184	Junior college / Vocational school	237
	University	362	University	433
	Graduate school	167	Graduate school	63
	Other	15	Other	10
	I do not know	4	I do not know	4
	I do not want to answer	5	I do not want to answer	16
Supporting political party	Republican Party	366	Liberal Democratic Party (自民党)	294
	Democratic Party	422	The Constitutional Democratic Party of Japan (立憲民主党)	42
	–	–	Democratic Party For the People (国民民主党)	10
	–	–	Komeito (公明党)	17
	–	–	Japanese Communist Party (日本共産党)	16
	–	–	Nippon Ishin (日本維新の会)	59
	–	–	Social Democratic Party (社民党)	6
	–	–	The Party to Protect the People from NHK (NHKから国民を守る党)	8
	–	–	Reiwa Shinsengumi (れいわ新選組)	15
	Other political party	43	Other political party	2
	I do not support any party	197	I do not support any party	501
	I do not know	23	I do not know	85
	I do not want to answer	12	I do not want to answer	53
Time of PC/smartphone use	Less than 30 min	10	Less than 30 min	17
	30 min to 1 h	37	30 min to 1 h	105
	1 h–2 h	98	1 h to 2 h	235
	2 h–3 h	156	2 h–3 h	245
	3 h–4 h	192	3 h–4 h	168
	4 h–6 h	187	4 h–6 h	162
	6 h–8 h	147	6 h–8 h	73
	8 h–10 h	122	8 h–10 h	37
	10 h or more	107	10 h or more	47
I do not use them at all	7	I do not use them at all	19	

Questionnaire sentences were the same in Japan and US. The answering options were slightly different in living area/states, house income and supporting political party

## Appendix 2. Scenarios (a)–(d) in English and Japanese

Scenario 1:

[English].

I am a researcher in the field of information science.

Using AI technology, I can analyze the voice and behavior of a deceased singer to virtually bring him or her back to life. With further research, we may be able to make the virtual singer sing new songs and perform as if he or she

were alive. At the same time, there is debate on the appropriateness of conducting business without receiving approval from the singer, while alive and some say that this project itself may harm the reputation of the singer.

However, this technology is already in practical use, and virtually revived singers are appearing on TV and new songs are being sold. In the future, I believe that singers from previous eras who have been revived through AI will be enjoyed and accepted by society.

May I continue on with this research?

[Japanese].

私は情報科学分野の研究者です。

AIの技術を使えば、亡くなった歌手の歌声や立ち振る舞いを分析して、バーチャルに甦らせることができます。この研究が進めば、あたかもその歌手が活着ているかのように新曲を歌わせ、パフォーマンスをさせることができるかもしれません。その一方で生前の歌手に許可を得ずにビジネスを行うことの是非や、そもそもこのような取り組み自体がその歌手の尊厳を傷つけるのではないかという議論もあります。

もっとも、すでにこの技術は実用化されており、バーチャルに甦った歌手がテレビに登場したりその新曲が発売されたりしています。私は将来、AIによって過去の歌手が再び活躍を始めることは社会に受け入れられ、人々を楽しませることにつながるかと考えています。

私はこの研究を進めても良いでしょうか?

Scenario 2:

[English].

I am a researcher in the field of information science.

Using AI technology, I can analyze data such as customer's purchase histories and keyword searches to predict that person's preferences. With further research, customers may be able to find products that they like by simply accessing a website. Meanwhile, some question the appropriateness of companies guiding the behavior and preferences of customers without their knowledge based on the company's intent.

However, many companies are already adopting this technology. In the future, I believe AI marketing will be further accepted by society and that it will become a methodology that will strongly support economic activities.

May I continue on with this research?

[Japanese].

私は情報科学分野の研究者です。

AIの技術を使えば、顧客の購買履歴や検索キーワードなどのデータを分析して、その嗜好を予測することができます。この研究が進めば、顧客はウェブサイトアクセスただけで自分好みの商品に出会うことができるかもしれません。その一方で、企業の思惑に基づいて顧客の行動や嗜好が知らず知らずのうちに誘導されることの是非も問われています。

もっとも、すでにこの技術は多くの企業によって採用されています。私は将来、AIによるマーケティングはより広く社会に受け入れられ、経済活動を強力にサポートする方法論になると考えています。

私はこの研究を進めても良いでしょうか?

Scenario 3:

[English].

I am a researcher in the field of information science.

Using AI technology, we can create various unmanned weapons that can act autonomously. With further research, we can remove human combatants from dangerous operations and reduce human casualties; we may even be able to attack with greater deadly force than with human combatants. At the same time, there is the ethical question of

AI killing people and also discussions regarding the legal liabilities in the event of a malfunction.

However, many unmanned weapons are already being deployed in actual battlefields. In the future, I believe the usefulness of unmanned weapons will be accepted by society and that their performance will serve national interests.

May I continue on with this research?

[Japanese].

私は情報科学分野の研究者です。

AIの技術を使えば、自律的に行動できるさまざまな無人攻撃兵器を作ることができます。この研究が進めば、人間の戦闘員を危険な任務から外すことで人的被害を減らすことや、人間の戦闘員よりも高い殺傷能力で攻撃することができるかもしれません。その一方で、人工知能が人間を殺すことの倫理的問題や、誤作動が生じた時の法的責任について議論が続いています。

もっとも、すでに実際の戦場には多くの無人攻撃兵器が投入されています。私は将来、これらの無人攻撃兵器の有用さが社会に受け入れられ、国益に叶う活躍をすると考えています。

私はこの研究を進めても良いでしょうか?

Scenario 4:

[English].

I am a researcher in the field of information science.

Using AI technology, if we analyze history of peoples' activities through their stored history in databases of personal credit cards, security cameras, etc., we may be able to predict possible criminal activities, such as theft and murder at a higher precision. With further research, we can prevent various crimes from occurring, and limit the activities of people that are more likely to commit a crime. Meanwhile, the appropriateness of using personal information for monitoring people and limiting the rights of people based on predictions is under debate.

However, information such as the activity history of an individual is already being analyzed by AI and used in marketing. In the future, I believe people will accept AI technology that goes further into personal privacy and that it will be of great benefit for protecting the safety and security of society.

May I continue on with this research?

[Japanese].

私は情報科学分野の研究者です。

AIの技術を使えば、個人のクレジットカードの使用履歴、防犯カメラ等から得られる行動履歴を分析することで、将来おきるであろう窃盗・殺人などの犯罪行為を高い精度で予測することができるかもしれません。この研究が進めば、さまざまな犯罪を未然に防ぐことや、犯罪を起こす可能性が高い人の行動を制限することができます。その一方で、個人情報の人々の監視のために使うことや、予測に基づいて人々の権利を制限することの是非も問われています。

もっとも、すでに個人の行動履歴等の情報はAIで分析され、マーケティングに活用されています。私は将来、個人のプライバシーに踏み込むAI技術も人々に受け入れられ、社会の安全・安心を守ることに大いに役立つと考えています。

私はこの研究を進めても良いでしょうか?

## Appendix 3

See Appendix Table 6

**Table 6** Item-total and item-remainder correlation coefficients. Pearson's correlation coefficient calculated by item-total correlations

	Total scores			
	Scenario (a)	Scenario (b)	Scenario (c)	Scenario (d)
Q1. Very fair/ Not fair at all	0.86	0.85	0.88	0.87
Q2. Makes a lot of sense/Does not make sense at all	0.86	0.80	0.86	0.82
Q3. Ethically very correct/incorrect	0.88	0.88	0.92	0.90
Q4. Positive impact to the future/ Negative impact to the future	0.85	0.85	0.91	0.86
Q5. Extremely favorable/unfavorable from a cultural perspective	0.86	0.86	0.91	0.89
Q6. Extremely favorable/unfavorable from a traditional perspective	0.86	0.84	0.89	0.88
Q7. Extremely favorable/unfavorable from a religious perspective	0.81	0.77	0.83	0.83
Q8. Extremely favorable/unfavorable from a public perspective	0.87	0.87	0.91	0.87
Q9. Explanation to society is sufficient/insufficient	0.87	0.87	0.91	0.87
Q10. Explanation to individuals is sufficient/insufficient	0.87	0.86	0.91	0.87
Q11. Transparency is sufficiently/insufficiently maintained	0.87	0.85	0.91	0.88
Q12. Policies and laws are sufficiently/insufficiently established	0.86	0.85	0.89	0.86

Pearson's correlation coefficient calculated by item-remainder correlations

	Total scores											
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Scenario (a)	0.56	0.52	0.59	0.51	0.52	0.58	0.63	0.53	0.59	0.60	0.61	0.70
Scenario (b)	0.58	0.54	0.60	0.56	0.57	0.60	0.64	0.56	0.63	0.63	0.68	0.71
Scenario (c)	0.57	0.53	0.58	0.56	0.54	0.58	0.64	0.52	0.62	0.64	0.66	0.70
Scenario (d)	0.60	0.55	0.62	0.51	0.59	0.66	0.66	0.55	0.66	0.68	0.69	0.74

**Acknowledgements** We would like to thank Dr. Arisa Ema and Ms. Haruka Maeda for their insightful comments on the scenario design. We also thank the University of Tokyo graduate students for their comments on the AI quiz. This work was supported by the World Premier International Research Center Initiative (WPI).

**Author contributions** TH, YI, HY and NT constructed the surveys; TH analyzed the data; TH, YI, HY and NT contributed the interpretation of results and discussion.

**Funding** This work was financially supported by KAKENHI Grant No. 20K14464, MEXT, Japan and by SECOM Science and Technology foundation.

**Data availability** The data set and code are not publicly available due to ongoing analysis but are available from the first author upon reasonable request.

## Declarations

**Conflict of interest** The authors do not have any conflicts of interest.

**Ethics approval and consent to participate** This study received approval from the Institutional Ethics Committee of the University

of Tokyo (No. 20–120). The company obtained consent from the participants.

**Consent for publication** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Adebayo J, Kagal L, Pentland A (2015) The hidden cost of efficiency: fairness and discrimination in predictive modeling. Bloomberg Data for Good Exchange Conference.



- Albarrán I, Molina JM, Manuel J, Gijón C (2020). Perception of Artificial Intelligence in Spain. International Telecommunications society (ITS) Online Event.
- Bebeau MJ (2002) The defining issues test and the four component model: contributions to professional education. *J Moral Educ* 31:271–295. <https://doi.org/10.1080/0305724022000008115>
- Borenstein J, Drake MJ, Kirkman R, Swann JL (2010) The engineering and science issues test (ESIT): a discipline-specific approach to assessing moral judgment. *Sci Eng Ethics* 16:387–407. <https://doi.org/10.1007/s11948-009-9148-z>
- Chaar BB (2009) Professional ethics in pharmacy practice: developing a psychometric measure of moral reasoning. *Pharm World Sci* 31:439–449. <https://doi.org/10.1007/s11096-009-9292-1>
- Crain WC (1985) Kohlberg's stages of moral development. In: Crain WC (ed) *Theories of development*. Prentice-Hall, New York, pp 118–136
- Elm DR, Weber J (1994) Measuring moral judgment: the moral judgment interview or the defining issues test? *J Bus Ethics* 13:341–355. <https://doi.org/10.1007/BF00871762>
- Ema A (2017) Ethically aligned design dialogue: A case practice of responsible research and innovation (in Japanese). *Journal of the Japanese Society for Artificial Intelligence*, 32: 694–700. [https://doi.org/10.11517/jjsai.32.5\\_694](https://doi.org/10.11517/jjsai.32.5_694)
- Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M (2020) Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication. <http://nrs.harvard.edu/urn-3:HUL.InstREpos:42160420>
- Funk C, Tyson A, Kennedy B, Johnson C (2020) Science and scientists held in high esteem across global publics. Pew research center.
- Future of Life Institute (2015) Autonomous weapons: An open letter from AI & robotics researchers. <https://futureoflife.org/open-letter-autonomous-weapons/> Accessed 18 October 21
- Goto T, Mizumachi E, Kudo M, Kano, K (2014) Verification on the Utility of Using the Segmentation Method Developed in Australia to Assess Audiences of Science and Technology Events (in Japanese). *Japanese Journal of Science Communication* 15: 17–35. <https://doi.org/10.14943/66441>
- Hagendorff T (2020) The ethics of AI ethics: an evaluation of guidelines. *Minds Mach* 30:99–120
- Haidt J (2012) *The righteous mind: Why good people are divided by politics and religion*. Pantheon Books, New York
- Haidt J, Joseph C (2004) Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus* 133:55–66
- Haidt J, Joseph C (2007) The moral mind: How 5 sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In: Carruthers P, Laurence S, Stich S (eds) *The innate mind*, vol 3. Oxford University Press, New York, pp 367–391
- Hosotsubo M, Tsunoda H, Kano K, Okumura A, Hoshino T. (2020) Public attitudes to science and technology: social acceptance of new technologies. National Institute of Science and Technology Policy (NISTEP), Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. <https://doi.org/10.15108/rm296> Accessed 18 October 2021
- Jensen LA (2011) The cultural development of three fundamental moral ethics: Autonomy, community, and divinity. *Zygon* 46:150–167. <https://doi.org/10.1111/j.1467-9744.2010.01163.x>
- Jensen LA (n.d.) Three ethics measures. <https://lennettjensen.com/three-ethics-measures/>. Accessed 18 October 2021
- Kohlberg L (1968) The child as a moral philosopher. *Psychol Today* 2:24–30
- Kruijtbosch M, Göttgens-Jansen W, Floor-Schreudering A, van Leeuwen E, Bouvy ML (2019) Moral reasoning among Dutch community pharmacists: testing the applicability of the Australian professional ethics in pharmacy test. *Int J Clin Pharm* 41:1323–1331. <https://doi.org/10.1007/s11096-019-00869-5>
- McGraw KM, Bloomfield J (1987) Social influence on group moral decisions: the interactive effects of moral reasoning and sex role orientation. *J Pers Soc Psycho* 53:1080–1087
- Murayama Y, Miura A (2019) Validation of the Japanese Version of the moral foundation questionnaire: Investigating the relationship with ideologies (in Japanese). *Jpn Psychol Res* 90:156–166. <https://doi.org/10.4992/jjpsy.90.17234>
- Ouchchy L, Coin A, Dubljević V (2020) AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. *AI Soc* 35:927–936. <https://doi.org/10.1007/s00146-020-00965-5>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
- PESTI (n.d.) What's Victorian segment. [https://www.nistep.go.jp/wp/wp-content/uploads/Infographics\\_VictorianSegments.pdf](https://www.nistep.go.jp/wp/wp-content/uploads/Infographics_VictorianSegments.pdf) Accessed 18 October 21
- Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1:81–106. <https://doi.org/10.1007/BF00116251>
- Reidenbach RE, Robin DP (1988) Some initial steps toward improving the measurement of ethical evaluations of marketing activities. *J Bus Ethics* 7:871–879. <https://doi.org/10.1007/BF00383050>
- Reidenbach RE, Robin DP (1990) Toward the development of a multi-dimensional scale for improving evaluations of business ethics. *J Bus Ethics* 9:639–653. <https://doi.org/10.1007/BF00383391>
- Rest J, Cooper D, Coder R, Nasanz J, Anderson D (1974) Judging the important issues in moral dilemmas: an objective measure of development. *Dev Psychol* 10:491–501
- Rest JR, Davison ML, Robbins S (1978) Age trends in judging moral issues: a review of cross-sectional, longitudinal, and sequential studies of the Defining Issues Test. *Child Dev* 49:263–279. <https://doi.org/10.2307/1128688>
- Rest JR, Narvaez D, Thoma SJ, Bebeau MJ (1999) DIT2: Devising and testing a revised instrument of moral judgment. *J Educ Psychol* 91:644–659. <https://doi.org/10.1037/0022-0663.91.4.644>
- Self DJ, Ellison EM (1998) Teaching engineering ethics: assessment of its influence on moral reasoning skills. *J Eng Educ* 87:29–34. <https://doi.org/10.1002/j.2168-9830.1998.tb00319.x>
- Shweder RA, Much NC, Mahapatra M, Park L (1997) The “big three” of morality (autonomy, community, and divinity), and the “big three” explanations of suffering. In: Brandt A, Rozin P (eds) *Morality and health*. Routledge, New York, pp 119–169
- Victorian Department of Innovation, Industry and Regional Development (2011) *Community Interest and Engagement with Science in Victoria*.
- World Values Survey Association (n.d.). World Values Survey Wave 7: 2017–2020. <https://www.worldvaluessurvey.org/WVSONline.jsp> Accessed 18 October 21
- Von Schomberg R (2011) Towards responsible research and innovation in the information and communication technologies and security technologies fields. European Commission. [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/mep-rapport-2011\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/mep-rapport-2011_en.pdf)
- Yamagishi A (1976) The development of moral judgement (in Japanese). *Japanese J Educ Psychol* 24:97–106. [https://doi.org/10.5926/jjep1953.24.2\\_97](https://doi.org/10.5926/jjep1953.24.2_97)
- Yamaha Corporation (n.d.) Misora Hibari VOCALOID:AD (in Japanese) [https://www.yamaha.com/ja/about/ai/vocaloid\\_ai/](https://www.yamaha.com/ja/about/ai/vocaloid_ai/) Accessed 18 October 21
- Zhang B, Dafoe A (2019) Artificial intelligence: American attitudes and trends. [https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/us\\_public\\_opinion\\_report\\_jan\\_2019.pdf](https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/us_public_opinion_report_jan_2019.pdf) Accessed 18 October 21

Zhang B, Dafoe A (2020) US public opinion on the governance of artificial intelligence. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society: 187–193.

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