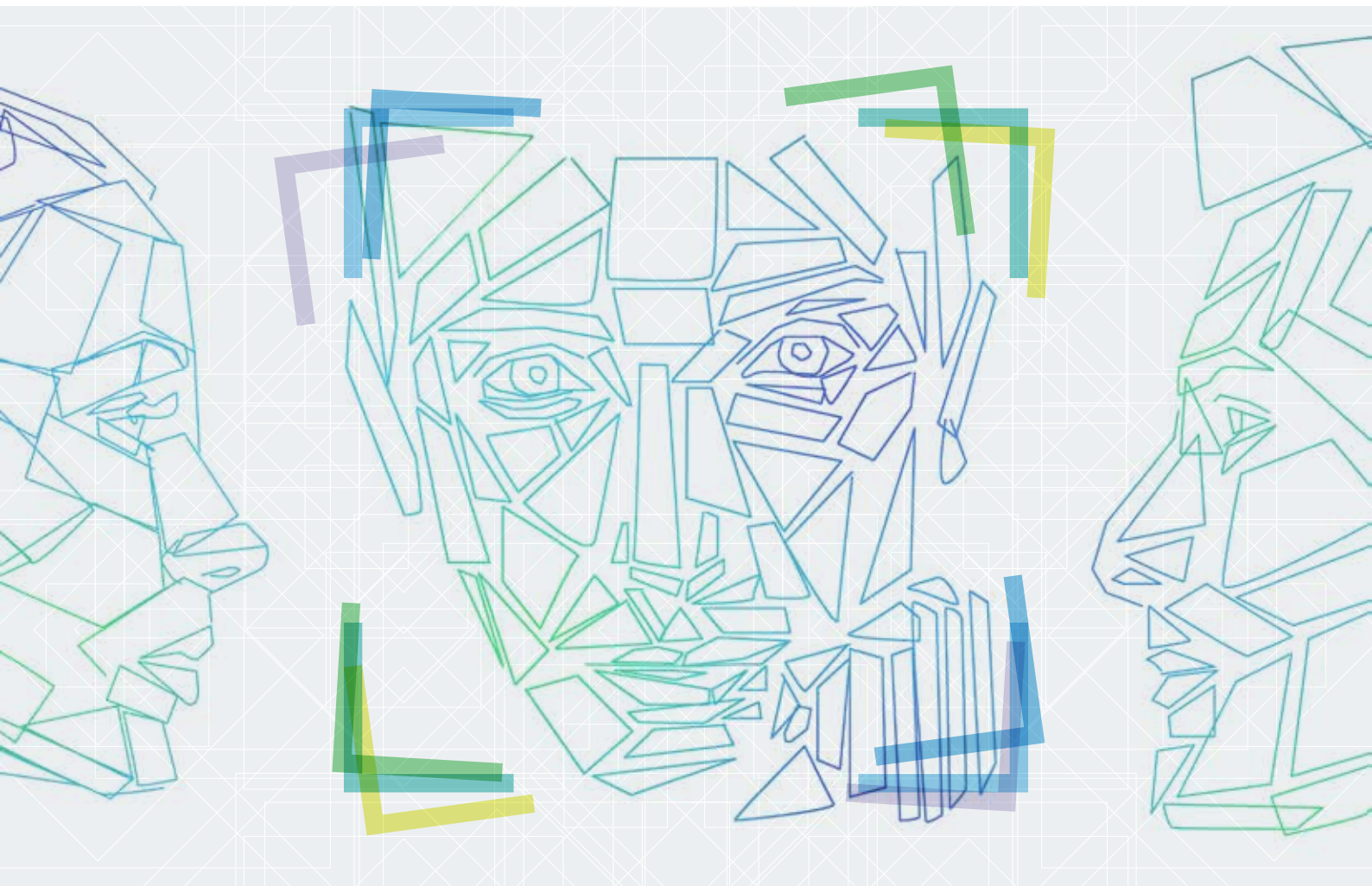




# Simulating a Multi Stakeholder Dialogue about Facial Recognition Technology: A Responsible Innovation Workshop for College Students



**Arisa Ema** The University of Tokyo / RIKEN

**Moe Sunami** Pomona College

**Shiori Komatsu** The University of Tokyo / Legal Apprentice Appointed by the Supreme Court of Japan (74th)

**Charles Ovink** United Nations Office for Disarmament Affairs

**Takashi Matsumoto** The University of Tokyo / Deloitte Touche Tohmatsu LLC

## Abstract:

Interdisciplinary, cross-sector collaboration is necessary to AI governance. To cultivate future leaders who can collaborate across sectors, and view issues from different perspectives, the UN Office of Disarmament Affairs and the University of Tokyo staged a workshop for undergraduate and graduate students in Japan that centered on an interactive, simulated multi stakeholder dialogue. Participants were asked to reach consensus on whether to permit commercial and police uses of facial recognition technology, and on what conditions, by taking the role of the police, government, tech company, or civil society. The interactive activity was preceded by introductory lectures on fairness and other issues by lawyers, computer scientists, policymakers, and social scientists. After the workshop, participants reported being more willing to participate in AI governance in the future because broad participation has potential to actually shape how AI is used in society. The hypothetical scenario was able to bring global conversations on police surveillance and race into a Japanese context, and prompted students to think flexibly about what conditions should be met before expanding uses of AI. The organizers hope that this workshop format and resource may be used for future workshops for broader audiences, including populations whose voices are underrepresented in AI development, students in other countries, and the private sector.



# Table of Contents

<b>1. Motivation: Fostering Interdisciplinary Cooperation and Developing Future Leaders in AI Governance</b>	<b>02</b>
<b>2. Background and Workshop Outline</b>	<b>03</b>
2.1 UNODA Workshops on Responsible Innovation	03
2.2 Workshop Structure and Theme	03
2.3 Event Participants	05
<b>3. Workshop Summaries</b>	<b>06</b>
3.1 Introductory Lectures	06
3.1.1. Introduction to Science, Technology, and Society (Ema)	06
3.1.2. "Challenges in Societal Implementation of AI" (on Fairness and Explainability) (Dr. Hiromi Arai)	06
3.1.3. "Responsible Innovation" (Ovink)	07
3.1.4. "Legal & Practical Issues in AI Implementation" (Mr. Naohiro Furukawa)	08
3.1.5. "AI and the Japanese Constitution" (Professor Tatsuhiko Yamamoto)	08
3.2. Q&A Session with Ms. Izumi Nakamitsu	09
3.3 Interactive Activity for Simulating Multi Stakeholder Dialogue	10
3.3.1 Hypothetical Scenario and Discussion Prompts	10
3.3.2 Summary of Student Discussions	11
3.3.3 Summary of Mock Multi-Stakeholder Discussion by Mentors	13
3.3.4 Comments from a Media Perspective (Mr. Ryohei Yasoshima, Nikkei Shimbun)	16
3.4 Post-Workshop Evaluation Results	16
3.4.1 Overall Satisfaction with the Workshop	16
3.4.2 Student Learnings	16
<b>4. Discussion and Future Directions</b>	<b>18</b>
4.1 Local Contexts and Global Debates: Trust, Policing, and Surveillance	18
4.2 Notes on the Framing of the Discussion	19
4.3 Future Directions	19
Authors' Profiles	21
Appendix 1: Workshop Scenario	22

# 1 Motivation: Fostering Interdisciplinary Cooperation and Developing Future Leaders in AI Governance

As safety and fairness emerge as salient issues in artificial intelligence (AI), there is an international effort to create and share principles and case studies [Jobin 19]. The AI research community has also begun to recognize the necessity of exploring Ethical, Legal and Social Implications (ELSI) of AI from the development stage, and as one example of this ongoing movement, the Ethics Committee of the Japanese Society for Artificial Intelligence have established the ELSI Award<sup>[1]</sup>. The term ELSI originates from its use within the Human Genome Project in 1990, where 3% to 5% of the total grant was to be spent on research on ethical, legal and social issues. In practice, however, the Genome Project's humanities and social science researchers were dominated by STEM researchers in setting the research agenda, and lacked the opportunity to provide meaningful critique.

Starting around 2010, under the banner of "ELSI 2.0" or "Responsible Research and Innovation" (RRI), calls for a more cooperative research effort among interdisciplinary researchers have gained momentum. Approximately a decade after, progress under RRI is being evaluated. It was found that forms of cooperation are diverse and defies generalization, in part due to tendencies of researchers to stick within their own disciplines. Within the field of AI, there exist multiple levels of cooperation, such as collaboration between research groups [Ema 20], government-led committees [Shiroyama 18a], or dialogues convened by international organizations such as the United Nations. The overall process by which different actors work in tandem to assess technology's impact, make decisions and policies, and implement technologies can be collectively referred to as "governance" of technology [Shiroyama 18b].

One aspect of this cooperation is diversity and inclusion [Ema 21]. These overlap with a number of Sustainable Development Goals (SDGs) described by the United Nations. As AI pervades through society, it is important to develop a future workforce who can understand AI's challenges and possibilities, communicate in an accessible manner to a range of stakeholders, and advance constructive dialogue. The University of Tokyo has heeded this challenge by starting, since 2017, to offer joint-discipline classes across information science, public policy, cultural studies, ethics, and the social sciences, consisting of lectures and group discussions<sup>[2]</sup>. Those classes were only available to enrolled students at University of Tokyo, motivating the authors to organize a workshop that is open to the public.

[Jobin 19] Anna Jobin, Marcello Lenca & Effy Vayena: The global landscape of AI ethics guidelines, *Nature Machine Intelligence*, 1, 389-99, 2019

[1] <http://ai-elsi.org/archives/793>



[Ema 20] Arisa Ema, *AI and Society: A Pathway from Interdisciplinary-alone to Interdisciplinary Research*, 『学術の動向』, vol. 25, No. 7, 2020, pp. 29-37. [https://doi.org/10.5363/tits.25.7\\_29](https://doi.org/10.5363/tits.25.7_29)

[Shiroyama 18a] Hideaki Shiroyama: *Jinko-Chinou to Technology Assessment: Wakugumi • Taisen to Jikkenteki Torikumi (AAI and Technology Assessment Framework, Institutions and Experimental Attempts)*, *Kagaku-Gijutsu Shakairon Kenkyu (Journal of science and technology studies)*, Volume 16, p.65-80, 201

[Shiroyama 18b] Hideaki Shiroyama: *Kagaku-Gijutsu to Seiji (Science and Technology in Politics)*, *Minerva Shobo*, 2018

[Ema 21] Arisa Ema: *Commentary: From AI Principles to Practice: Lessons for Japan to Learn from International Activities, AI Principles to Practice*, pp. 39-42, <https://ifi.u-tokyo.ac.jp/en/projects/future-ai/>

[2] [http://science-interpreter.c.u-tokyo.ac.jp/ai\\_society/](http://science-interpreter.c.u-tokyo.ac.jp/ai_society/)

## 2 Background and Workshop Outline

---

### 2.1 UNODA Workshops on Responsible Innovation

The United Nations Office for Disarmament Affairs (UNODA) has been organizing workshops on “Responsible Innovation” for STEM students across the world, motivated by the UN Secretary General’s call to “encourage responsible innovation by industry, engineers, and scientists” under the Agenda for Disarmament. AI research fields to date have largely ignored peace and security dynamics, so AI researchers can often be unaware of the potential negative implications of their work. Charles Ovink, a UNODA Political Affairs Officer and a co-author, spoke with Ema at a UN-related conference in July 2020 to initiate an effort to hold a workshop in Japan. The Ethics Committee of the Japanese Society for Artificial Intelligence and Next Generation Artificial Intelligence Research Center, the University of Tokyo joined in supporting the workshop held in Japan.



### 2.2 Workshop Structure and Theme

The workshop took place over two four-hour blocks in the morning on February 8th and 9th 2021 in Japan Standard Time, so as to accommodate an online format and time differences across Japan and the United States where the UN is headquartered. Day 1 consisted of lectures on fairness in data science and artificial intelligence, and Day 2 consisted of an interactive discussion-based activity where participants were prompted to assess the costs and benefits of an application of AI from different perspectives, and practice proposing a path forward through discussion (Table 1). An additional aim of the interactive portion was to encourage participants to move flexibly across the public sector, private sector, and academia. Mentors for the interactive activity provided participants with readings and videos to review between Day 1 and Day 2, intended to take about 20 minutes of preparation.

The interactive activity used two configurations for discussion: for a part of the workshop, participants gathered by their assigned stakeholder role, and discussed questions with an assigned mentor with expertise in the role. For other parts of the workshop, participants gathered with representatives from other stakeholders (“jigsaw format”) without the guidance of a mentor. The theme of the discussion was a fictional scenario<sup>[3]</sup> concerning the use of facial recognition technology (FRT) in stores and by the police. The stakeholder roles consisted of the IT company contributing FRT, the government, civil society, and the police (Figure 1).

[3] See <https://sites.google.com/view/aiworkshop-ut/scenario>

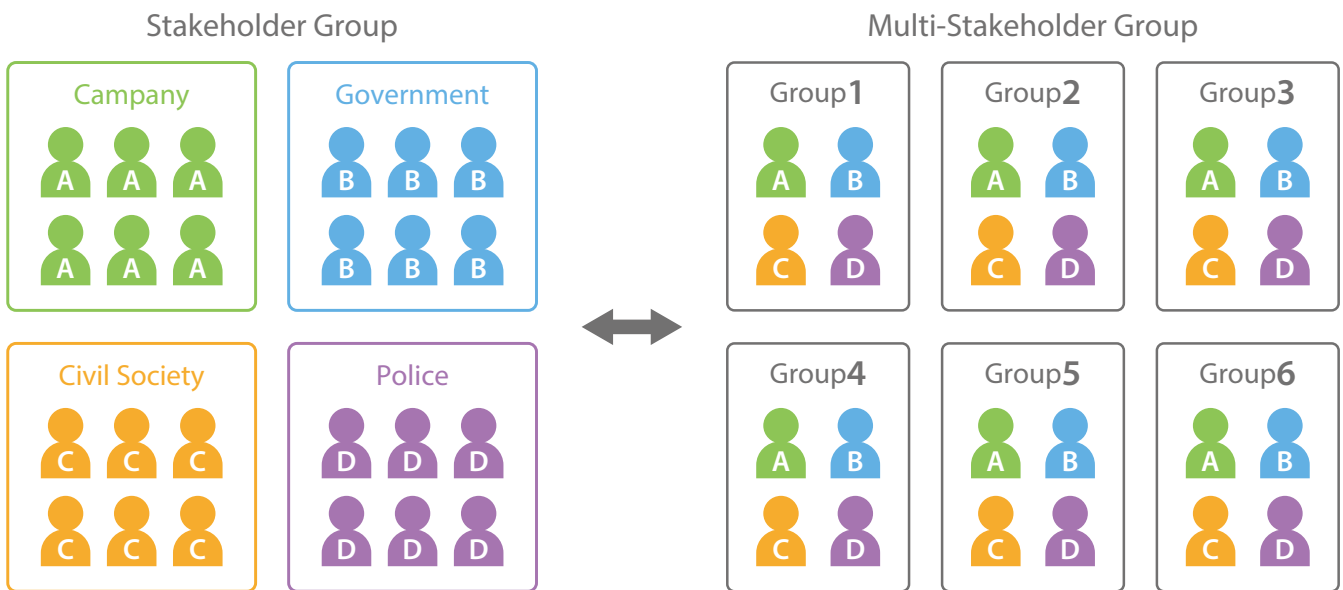


Figure 1 : Jigsaw-style" discussion configuration

FRT was chosen as a theme for several reasons. FRT is used in retail stores and airports in Japan, so it was considered an accessible issue. Additionally, the authors conducted a preliminary survey<sup>[4]</sup> in Fall 2020 to gauge public interest in different themes discussed in AI ethics. The survey described scenarios for different societal issues brought about by AI (privacy, fairness, misuse and misinformation, safety, human dignity and norms, and future of work) and asked respondents to rate their importance and relevance to students. Notably, there appeared to be generational differences on which issues respondents found significant, where respondents under the age of 25 displayed more interest in fairness issues. FRT was then selected because its applications often have consequences for fairness. Furthermore, from a responsible innovation perspective, it is important to release the content of multi stakeholder discussions to a wide audience. For this reason, the organizers invited a member of the media to the interactive discussions on Day 2 who could advise participants on how their discussions would be featured on the media, and how to talk with the media (Table 1). Finally, participants were given a survey at the end of the workshop to evaluate their awareness of relevant issues and the program's educational goals.

[4] Survey content and results can be seen on <https://sites.google.com/view/aiworkshop-ut/survey>



Table 1 : Summary of the two-day program

<b>Day 1</b> (February 8)	<p>9:00-9:15 Opening remarks by Charles Ovink (Political Officer, UN Department for Disarmament Affairs), Arisa Ema (Project Assistant Professor, Institute for Future Initiatives at the University of Tokyo)</p> <p>9:15-9:40 "Introduction to Science, Technology and Society," Arisa Ema</p> <p>9:40-10:40 "Challenges in Societal Implementation of AI," Dr. Hiromi Arai (Unit Leader, AI Safety and Reliability Unit, RIKEN Center for Advanced Intelligence Project)</p> <p>10:40-10:45 Break</p> <p>10:45-11:15 "Responsible Innovation," Charles Ovink</p> <p>11:15-12:15 "Legal and Practical Issues in AI Implementation," Naohiro Furukawa (Lawyer, ABEJA Corporation)</p> <p>12:15-12:20 Break</p> <p>12:20-12:30 Explanation of Day 2's activities</p> <p>12:30-13:00 "AI and the Constitution," Professor Tatsuhiko Yamamoto (Keio Law School)</p>
<b>Day2</b> (February 9)	<p>9:00-10:00 Conversation with Ms. Izumi Nakamitsu (UN Under-Secretary-General, High Representative for Disarmament Affairs)</p> <p>10:00-11:40 Interactive activity on multi-stakeholder dialogue, with mentors</p> <ul style="list-style-type: none"> <li>• Mr. Yoichi Iida (Director, Information and Communication Policy Research, International Strategy Bureau, Ministry of Internal Affairs and Communications)</li> <li>• Dr. Atsuko Sano (Project researcher, Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Part-time Lecturer, Graduate School of Social Design Studies, Rikkyo University)</li> <li>• Mr. Kazuhiro Tsubobara, Senior Professor, Police Policy Research Center, National Police Academy</li> <li>• Mr. Naohiro Furukawa (ABEJA Corporation)</li> </ul> <p>11:40-11:45 Break</p> <p>11:45-12:20 Participant presentations</p> <p>12:20-12:25 Break</p> <p>12:25-12:50 Mock multi stakeholder dialogue and comments by mentors</p> <p>12:50-12:55 Comments on the media, Mr. Ryohei Yasoshima (Digital Policy Editor, Nikkei Shimbun Editorial Bureau)</p> <p>12:55-13:00 Closing Remarks</p>

## 2.3 Event Participants

In the interest of facilitating small group discussions, workshop capacity was limited to 36 participants, and in the end, 41 participants attended the event. Participant age ranged from 19 to 52 years old, and their academic institutions included, in addition to University of Tokyo, other domestic institutions such as Keio University, Waseda University, Hitotsubashi University, Meiji University, Tsuda College, and United Nations University, as well as universities abroad including Columbia University, Wesleyan University, Smith College, and University of Toronto, for a total of over 20 institutions. As for gender distribution, 29 women and 12 men attended. Participants had a wide variety of interests, including immigration and refugee issues, international relations, human rights, gender, national security, programming, math, law, big data, and agriculture. Participants who attended all sessions of the workshop were presented with certificates of completion from the University of Tokyo and UNODA.



# 3 Workshop Summaries

---

## 3.1 Introductory Lectures

### 3.1.1. Introduction to Science, Technology, and Society (Ema)

Ema, one of the workshop organizers, kicked off the workshop by introducing the discipline of Science, Technology, and Society Studies (STS), and explained how STS has contributed ways of thinking that were relevant to the learning objectives of the workshop. The term “STS” refers to both research in the theories of science and technology (“Science and Technology Studies”), as well as praxis at the intersections of science, technology, and society through communication and policy formation. She emphasized that issues concerning AI include problems that cannot be solved only by the tools of science (also referred to as “Trans-Science” [Weinberg 72]), and that it is important to advance both research and practice in tandem. She framed the objectives of the workshop as encouraging participants to notice their hidden assumptions and “blind spots,” and to approach problems without clear answers from different perspectives.

Ema also brought up the concept of Collingridge’s dilemma, which states that while it is difficult to predict the influence of technology before it is used in society, it is hard to control technology once it becomes widespread. Ema pointed out that it is important to think of artificial intelligence technology not in terms of a “social experiment,” where the technology is used experimentally by certain people, but in terms of an “experimental society” in which we ourselves are a part. Finally, she concluded by saying that in such an “experimental society,” it is important for every one of its residents, not only researchers and experts, to take on responsibility for how we respond to new technologies and evolving social values, and that this attitude will promote responsible innovation.

### 3.1.2. “Challenges in Societal Implementation of AI” (on Fairness and Explainability) (Dr. Hiromi Arai)

Next, Dr. Hiromi Arai (Unit Leader, AI Safety and Reliability Unit, RIKEN Center for Advanced Intelligence Project) introduced students to concepts of fairness and explainability in the context of ethical AI. In addition to the opacity of algorithms (“black boxes”) and potential violations of privacy, there are ethical concerns about the discriminatory outcome of machine learning algorithms and predictive models. It has been found that a model used to predict recidivism, COMPAS<sup>[5]</sup>, assigned a higher likelihood to Black defendants than to white defendants. In another example, Amazon halted its use of its hiring AI after it was found to make discriminatory judgements against applicants who were women<sup>[6]</sup>.

Ms. Arai explained that bias in AI can have causes in the generation and collection of data, such as over- or under-representation in the learning dataset, and in the learning model itself which may, for instance, fail to account for outliers



**Arisa Ema**

Project Assistant Professor, Institute for Future Initiatives at the University of Tokyo

---

[Weinberg 72] Alvin, M. Weinberg, Science and Trans-Science, Minerva, vol 10, No.2, p.209-222, 1972.



**Hiromi Arai**

Unit Leader, AI Safety and Reliability Unit, RIKEN Center for Advanced Intelligence Project

---

[5] <https://globe.asahi.com/article/12287549>

[6] <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>



and minority samples. She noted that efforts to combat unfairness in machine learning include alternative learning models or the use of benchmarks during audits. Additional difficulties remain in pursuing fairness, including choosing the appropriate fairness criteria and balancing trade-offs with accuracy.

Explainability of machine learning was introduced as one method gaining a better understanding of how AI makes its decisions and, consequently, what AI should be used for. There are two methods of explanation: model explanation, which explains a complex model with an interpretable model, and outcome explanation, which explains the output of a black-box model for a specific input. While some explanatory techniques exist, Dr. Arai noted that there is room for further research, and warned against “fair-washing” (when an algorithm is presented as though it meets certain ethical values when it does not).

### 3.1.3. “Responsible Innovation” (Ovink)

Ovink examined what is meant by “responsible innovation,” which is currently an urgent issue for disarmament, peace and security. He noted that the impact of technical decisions are often unclear, and are made with limited transparency and engagement with society at large. In order to appropriately deal with the impact of complex scientific and technological systems, the cooperation of many stakeholders — not only between countries, academics, researchers, and industries, but also between individual citizens — is necessary.

One global-scale problem in military or weaponized use of AI is that it can increase unpredictability, while another issue is the relatively slow pace of international agreements to manage peace and security implications of new technologies. Looking at the example of facial recognition technology and the use of laser instruments to disrupt equipment, Ovink noted that these uses are not covered under the Convention on the Prohibition and Restriction of the Use of Certain Conventional Weapons (CCW) unless they are used against humans.

Using an online translation tool as an example, Ovink illustrated some issues of the unintended replication and exacerbation of bias due to a biased dataset. The dataset is extracted from a gender-biased society, and the training results reproduce inherent biases within society. He also pointed out that the source of information is a serious issue in the context of military or weaponized use of AI, including autonomous lethal weapon systems (LAWS), which is one of the areas of interest for the UN Office for Disarmament Affairs.

Finally, he reiterated that in today’s global society, where decisions and legislation of one country can affect other countries, cooperation of all stakeholders in industry, academia, government, and the private sector is essential. Citing an earlier example that it is difficult to determine who should be held legally responsible for an accident involving the use of a self-driving car<sup>[7]</sup>, he touched on the difficulty of creating agreement and regulations based on international consensus.



**Charles Ovink**

Political Officer, UN Office for Disarmament Affairs

[7] <https://www.bbc.com/news/business-43459156>

### 3.1.4. “Legal & Practical Issues in AI Implementation” (Mr. Naohiro Furukawa)

Mr. Furukawa (Attorney at Law, ABEJA Corporation) introduced the current status of regulations surrounding FRT, and commented on the relationship between regulation and innovation.

In the United States, a number of FRT regulations have been enacted since 2019. As of late, 2020 and 2021 saw not only regulations that banned the use of FRT by the public sector in some cities<sup>[8]</sup> and states<sup>[9]</sup>, but also the use of FRT in public spaces and in schools, indicating that authorities are treating FRT with caution. Meanwhile, in Japan, there is no law that directly regulates FRT, and the only applicable binding law is the Act on the Protection of Personal Information. Since January 2017 the Ministry of Economy, Trade and Industry (METI) and the Ministry of Internal Affairs and Communications (MIC) have published the nonbinding “Guidebook for Utilization of Camera Images,” which stipulates how businesses should build trust with consumers. While non-binding guidance is unenforceable, Mr. Furukawa found significance in how “soft” (non-binding) law can respond rapidly to expert input and to changes in society, can influence business behavior, and is available to be referred to by courts. Mr. Furukawa recommended that a combination of binding and non-binding laws be tailored to meet specific policy needs. From the perspective of promoting innovation, strict regulations can stifle innovation and allow only large companies to comply, but on the other hand, unregulated technologies will not be accepted by users and society, so rules with the appropriate strength are necessary. For start-ups with limited resources, Mr. Furukawa noted that, even given the existence of regulations on AI, they must figure out how they prioritize different ethical and market values.

Mr. Furukawa proposed that the difference in accuracy of FRT itself is less of an issue than the impact of the difference, according to each use case. Mr. Furukawa concluded by pointing out the following factors to take into context for each use case of fair AI: regional context, target group, definition of fairness, availability of data, and tradeoffs between fairness and accuracy.

Finally, Mr. Furukawa identified the difficulty of establishing communication among different stakeholders who do not share common knowledge, even though such close cooperation is important for working towards fairness in AI.

### 3.1.5. “AI and the Japanese Constitution” (Professor Tatsuhiko Yamamoto)

Mr. Yamamoto (Professor, Keio Law School) pointed out that the negative impacts of AI on the basic principles of the Constitution include the problems of privacy violation, violation of the right to self-determination, and reproduction of discrimination due to profiling. Recent, prominent cases of profiling by AI include the Cambridge Analytica case in the United States in 2016<sup>[10]</sup>, and the Rikunabi case in Japan in 2019<sup>[11]</sup>. Both of these cases attempted to identify and score individuals’ personalities and inner processes through their social network data and web browsing history, but Yamamoto believes that there is a danger that the increased accuracy of profiling will expose the instinctive side of individuals



**Naohiro Furukawa**

Lawyer, ABEJA Corporation

[8] <https://www.portlandmaps.com/bps/mapapp/proposals/#/private-facial>

[9] <https://www.governor.ny.gov/news/governor-cuomo-signs-legislation-suspending-use-and-directing-study-facial-recognition>



**Tatsuhiko Yamamoto**

Professor, Keio Law School

[10] <https://www.bbc.com/news/technology-43465968>

[11] <https://www.nikkei.com/article/DGXMZO52948640U9A201C110000/>

underneath the “masks” (personas) they wear.

Additionally, he warned that the right to self-determination that underpins the Japanese Constitution may be eroded under the attention economy (the concept that attention is becoming more scarce under the spread of the Internet and becomes a valuable commodity). Psychological profiling by AI can analyze a user’s cognitive tendencies, and send only the information to which the user is most responsive, so that others’ decisions replace self-determination.

Mr. Furukawa referenced Amazon’s hiring AI<sup>[12]</sup> and HireVue<sup>[13]</sup> to explain how biased datasets can exacerbate pre-existing structural inequalities. Moreover, if use of a specific scoring algorithm dominates the market and expands through society, this could marginalize individuals with low scores and keep them there in what he referred to as a “virtual slum”, especially if the scoring is opaque and there is no clear way to increase people’s scores.

Finally, Mr. Yamamoto touched on overseas trends in data protection. He explained that the EU General Data Protection Regulation was grounded in data protection as a basic human right, citing Article 22 (prohibition of fully automated decision-making) and Articles 13 to 15 (obligation to provide information). The United States, which has traditionally taken the view that regulating the use of data infringes upon freedoms, has also been focusing on data protection after the Cambridge Analytica scandal in 2016. Mr. Yamamoto pointed out that Japan, in contrast, has yet to take a position on the right to self-determination as it relates to information technology.

### 3.2. Q&A Session with Ms. Izumi Nakamitsu

Day 2 began with a moderated Q&A session with Ms. Izumi Nakamitsu (UN Under-Secretary-General and High Representative for Disarmament Affairs). In the introductory remarks, Ms. Nakamitsu emphasized that understanding the impacts of an accelerating development of science and technology, and advancing responsible innovation, is a high priority for the UN. From an international security perspective, she identified the need to develop codes of conduct for researchers regarding dual use technologies, especially to curb the proliferation of low-budget, dual use technologies that can be accessible to anyone over the Internet. Additionally, she brought up lethal autonomous weapon systems (LAWS) as another area where law has yet to catch up.

Participants brought a range of questions regarding the use of AI, control of dual use and general use technologies, and the role of states and the UN. In her responses, Ms. Nakamitsu commented that the United Nations is actively working on the positive use of AI, and there even exists an innovation hub within the UN. She explained that a debate is ongoing on how to ensure human oversight over AI, especially when it is related to the use of force, and that this debate is necessary given that the quality of decisions made by AI are limited by the characteristics of the datasets it is trained on.

On the topic of holding control over AI, Ms. Nakamitsu commented that although

[12] <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

[13] <https://www.technologyreview.com/2019/11/07/75194/hirevue-ai-automated-hiring-discrimination-ftc-epic-bias/>



**Izumi Nakamitsu**

UN Under-Secretary-General, High Representative for Disarmament Affairs

it is difficult to control the technologies themselves, approaches such as prohibiting the development of certain subsets of weaponized uses of AI and creating norms and frameworks to constrain the actions of states may be feasible. She responded that the UN is expected to play a role of thought leadership in setting up a framework that allows multiple stakeholders to participate in discussions side by side with governments, rather than limiting participation to representatives of governments. Ms. Nakamitsu concluded the session by stating that, in the age of AI, it is important to examine human values that make humans who they are.



Figure 2 : Group photo with Ms. Nakamitsu (upper right)

### 3.3 Interactive Activity for Simulating Multi Stakeholder Dialogue

#### 3.3.1 Hypothetical Scenario and Discussion Prompts

Students were assigned one of four roles: a tech company, the government, police, and civil society. The last role represented organizations that advocate for minority rights, in particular those of women and foreigners, whose issues have been receiving wider recognition in Japan.

For this exercise, the authors created a fictional scenario where FRT is applied to business and law enforcement contexts (Appendix 1). In the business scenario, a Japanese IT company “X” provides private retail facilities with FRT systems that finds customers displaying suspicious behaviors for theft, and reports them to store employees. This system has an additional feature to award redeemable “Eco-points” to registered customers who display environmentally conscious behavior, such as bringing a reusable bag or cup and purchasing environmentally friendly products.

In the law enforcement context, the authors proposed a setting where acts of terrorism with unidentified perpetrators have been occurring in public spaces frequented by many people. In response, the government has been bolstering

anti-terrorism efforts and requested the same company X to build an additional feature to its FRT systems to detect behaviors and people who may be linked to terrorism. The content of the algorithm has not been disclosed to the public, and it is unclear what characteristics would be singled out by the system, but the government is considering making this system mandatory among large retail facilities, and require sharing its results in real time with the police.

Both of these use cases in retail and law enforcement present trade-offs for participants to evaluate. These FRT systems can anticipate benefits for customer experience, the environment, theft prevention, and public safety. Yet there may be consequences and concerns for privacy, misidentification, public trust, and discrimination. There also must be a balancing of regulatory oversight and freedom to innovate, as well as a trade-off between accuracy and explainability of algorithms. Some trade offs require additional information to evaluate, such as whether a counter-terrorism AI is successful in identifying suspects.

Currently, similar scenarios have been advancing in Japan and worldwide. In 2019, it was reported that many bookstores in Tokyo have introduced facial recognition systems to prevent theft by sharing facial information of customers who have been confirmed as shoplifters<sup>[14]</sup>. Holding these actual contexts in mind, participants in the multi stakeholder activity were asked to reach a consensus on the following three questions, and discuss the conditions for agreement:

**Q1: Company X's facial recognition system should continue to be used commercially in the future. Yes / No**

**Q2: A counter-terrorism feature should be added to Company X's facial recognition system. Yes / No**

**Q3: As part of digital transformation (DX), the public and private sectors should be encouraged to use AI to inform their decision-making. Yes / No**

Students discussed the relationships between various stakeholders, and the interests, pain points, and constraints of the stakeholder positions they represented. They alternated between speaking amongst their own stakeholder group (guided by mentors) and multi-stakeholder discussions (just students), running through each configuration twice.

### 3.3.2 Summary of Student Discussions

The results of the students' discussions are shown in Table 2. Most groups agreed that FRT should not be banned outright in either retail or law enforcement. Students proposed a wide spectrum of conditions for FRT implementation, including binding and non-binding rules for privacy, accountability, security, accountability, transparency, fairness, scientific integrity, and human oversight.

Regarding Q1 on the use of FRT in retail facilities, most groups emphasized the need to ensure consumer choice through requirements for opt-in or opt-out practices, and displaying clear notice of intended use of data.

[14] <https://www.nikkei.com/article/DGXMZO47581150Q9A720C1SHJ000/> It should be noted that the information on the alleged shoplifters are not shared with the police, as the information is collected for harm prevention purposes.

Regarding Q2 on the use of FRT by law enforcement, discussion grew heated as concerns were raised about the accuracy of the system. Given the lack of clear characteristics to distinguish suspects, most groups recommended disclosing the basis on which the system identifies suspects and conducting third-party

Table 2: Participant answers to workshop questions

	Question 1.	Question 2.	Question 3.
Group 1	<b>Yes.</b> Ensure transparency, collaborate with police and public agencies with specialized knowledge regarding crime.	<b>No.</b> Dataset for terrorism suspects lacking. Datasets collected for anti-theft measures may target people of lower socioeconomic backgrounds. Private companies could retain the technology, but datasets related to terrorism should be managed or owned by the government.	<b>Yes.</b> Proceed cautiously while debating risks, and promote human values.
Group 2	<b>Yes.</b> Institute third party auditing, certification mechanisms.	<b>Partial Yes.</b> Limit usage, i.e., only use it for arresting suspects who have been identified, not stop-and-frisk. Ensure transparency & institute third-party audits and certification.	<b>Yes.</b> Decision-making should not be completely automated. Explain the technology clearly to drive acceptance by society.
	Private sector should have the lead in commercial use cases, while the government manages national security use cases. There should be guidelines at the national and industry level on personal data protection and dataset bias, provided there is space for competition. There should be clear notice of the purpose of data collection and the decisions it can impact. No decision should be completely automated and all results are subject to human review.		
Group 3	<b>Yes.</b> Create guidelines. Companies may be required to disclose algorithms as necessary. Users should have a choice on how their information is used.	<b>Yes,</b> only if the effectiveness and accuracy of the system is verified. Unclear what information is indicative of suspects. Require consent as a general rule and, in the event of acquiring information without consent, define clear legal conditions.	<b>Yes,</b> on a case-by-case basis. Create guidelines for what data should be collected. Disclose use purposes, data processors, and provide updates when use expands. Obtain consent.
Group 4	<b>Yes.</b> Allow users to opt in or opt out. Companies should create guidelines according to user demands. Enable anti-theft features, but cautious about "Eco points" for business incentives.	<b>Yes,</b> given there are binding laws governing its use and deployed only in emergency situations to save lives, similar to binding COVID-19 emergency orders. Should not lead to search of personal belongings.	<b>Yes.</b> The government can identify areas in which AI can be used, then respond according to the needs of various actors in industry, academia, and the public as public dialogue advances.
Group 5	<b>Yes.</b> Reinforce cybersecurity to prevent leaks. It may be difficult to obtain consent or give choice to users each time, so ensure transparency by disclosing some metrics for evaluation & having a third-party audit on fairness, explain results to users & gain trust. Limit use purposes to anti-theft.	<b>Lean No.</b> If in use, the government should install a division to implement this system. Ensure accuracy & explainability, verify with experiments, be clear about which variables affect results, conduct audits on effectiveness. Police would create standards for who is deemed suspect.	<b>Yes.</b> Humans oversight and input in decision-making necessary. Government should create guidelines on AI Ethics, and businesses would create self-regulating rules. Test effectiveness through pilots, ensure fairness, and conduct repeated audits to adjust standards iteratively.

Group 6	Yes. Use alternatives that do not use facial recognition technology when it is not necessary. Allow for rectification when someone is inaccurately deemed suspect. Make the purpose of use clear.	Yes. The government & IT companies should work closely together, especially on information management, and dispatch specialists in between. Clearly divide of responsibilities, i.e., IT companies store data, the police use the system when necessary. Limit installation to densely populated locations. Provide notice of data collection.	Yes. Progress cannot be held back. Humans should make the final decision and have responsibility to explain decisions.
Group 7	Yes. Obtain consent and notice for image capture and facial recognition, i.e., through in-store posters.	Yes. The government should clearly display the purpose of use, how and where data would be shared, stored, and managed. Conduct a third-party audit.	Yes. The influence of AI on decision-making should be proportional to the size of risk ("for criminal sentencing, AI should not have more than an 8% stake").
Group 8	Yes. Ensure explainability and transparency. Respect the right of customers to choose. The government should institute binding mechanisms for accountability and require third-party audits.	Yes. Companies have an obligation to disclose purposes of use and processing methods, assure cybersecurity. Government should create binding regulations restricting where the system can be used, disclosing standards used by the algorithm, and requiring explainability. Third party audits should be required.	Yes. Create internal guidelines within industries, companies. The government should legally define the minimum rights assured to users when artificial intelligence is used, which can set the foundation for police use as well. Final decisions should be made by humans.

audits to test effectiveness. In addition, groups called for the government to play a greater role, including taking responsibility for collecting and managing terrorism-related data and limiting the scope and parameters of the AI.

Finally, regarding Q3 on the use of AI in decision-making, groups broadly agreed that the spread of AI cannot be curtailed, but that humans should have final decision-making power. Additionally, when AI is used to make decisions in a law enforcement context, including for conducting search warrants or arrests, the role of AI in decisions should depend on its proven effectiveness.

### 3.3.3 Summary of Mock Multi-Stakeholder Discussion by Mentors

After student presentations, mentors held a mock discussion moderated by Ema. The mentors commented in their own personal capacity, while reflecting the general views of the stakeholders they represented. Main points are described in Table 3.

From a tech company perspective, Mr. Furukawa stated that tech companies should not introduce counterterrorism features to its FRT system. He emphasized that it was not the responsibility of companies to determine characteristics of terrorists, and that it would be far too large of a responsibility. National security was the government's responsibility and expertise, and outsourcing this process to different private companies risks creating fragmented, disparate standards.

From a police perspective, Mr. Tsubobara emphasized that human rights must be taken into consideration when private companies use AI to prevent crime.



**Kazuhiro Tsubobara**

Senior Professor, Police Policy Research Center,  
National Police Academy

Table 3: Responses to the simulated multi-stakeholder discussion by mentors

	Question 1. Should Company X be allowed to implement private uses of facial recognition technology in order to facilitate environmentally sustainable actions?	Question 2. Should the government use facial recognition technology in order to identify potential terrorist suspects?	Question 3. As part of digital transformation, we should encourage uses of AI to inform human decision-making (by private employees or by public authorities)
Company X	<b>Yes.</b> Companies benefit by reaching a bigger audience and increasing engagement. Customers benefit by gaining redeemable points. Employees benefit from automating tasks and reducing overwork. Security officers should get the final say on screening people.	<b>Lean No.</b> Private companies that aren't trained for counter terrorism should not be responsible for collecting data and developing algorithms that could sway national security. National government or international organizations should provide a standardized dataset and clear guidelines of what to look for, i.e., an official mugshot. Disjoint standards by different companies leads to confusion.	<b>Yes.</b> Progress cannot be held back. Conditions and risks should be discussed among stakeholders for different use cases. These would lead to non-binding laws or guidelines, and maybe resulting in binding laws.
Police	<b>Yes,</b> with reservations. Because the content of the algorithm is unclear, it's unclear whether the algorithm can be used to authorize police action. There must be transparency regarding what features are considered suspicious to build trust between the public and the police. Suspicious traits may vary by store type. Employees using the system should be trained well. Audit will ensure accountability.	<b>Yes,</b> so long as its potential discriminatory effects do not risk public trust. Shows promise as a scalable system. Must be cautious & transparent determining what suspicious traits are. Racial or cultural discrimination would lead to breakdown of trust between police and the public, hampering public safety & gathering necessary information to fight all types of crime.	<b>Yes.</b> Difficult to hold back progress. Should encourage people from all backgrounds and lifestyles to join discussion, debate, and incorporate their opinions into whether a technology should be adopted or not. In either case, each person should consider how they are affected by the adoption of new technologies, and identify pros and cons in relation to their own life, instead of relying on the positions of the government, academia, media, or companies.
Government	<b>Yes.</b> Ensure company's freedom to meet business demand and ensure customer's freedom to choose services. If Company X has a commercial demand for the service, they should be allowed to use it given they comply with regulations. Customers should make informed decisions and be able to opt out of the service or choose another store.	<b>Yes,</b> if it is possible to effectively identify potential terrorists. Dataset should be able to accurately identify which traits are indicative of terrorists, but this data may not exist. It should be made clear what AI can authorize police to do, and define those conditions based on accuracy and other factors.	<b>Yes.</b> Work on each use case one by one and ensure transparency. The best and fastest path to digital transformation is for diverse people to gradually build consensus on what is acceptable and what isn't, and for people to act autonomously accordingly. Laws should be considered the last resort.
Human rights / Residents	<b>Yes.</b> Both customers and the environment benefit from eco points. Customers should be notified about the use of facial recognition, how it is linked with their personal information, and its purpose. They should be able to opt out and have the right to be forgotten and their data erased when the data is no longer applicable.	<b>Yes,</b> but close to No. If an anti-terrorist capability is built on top of the existing system, the purpose of the data would differ. Additionally, concerned that the existing system's data would be biased along region, economic opportunity, and other factors, with unknown consequences if this data is used for national security purposes.	<b>No</b> for AI use by the police and country. Conditional yes for private companies. Japan should prioritize making more robust datasets on gender and foreigners, and other traits that can provide detail on social inequality.



Employees should be provided adequate guidance on how to use results gained from AI, and the police should also understand the inner workings of AI to determine whether it can be used as evidence.

Additionally, Mr. Tsubohara stressed the utmost importance for the police to maintain the public's trust in them as professionals, for the purposes of maintaining safety in society. Should the police rely on subpar AI, and engage in discriminatory behavior or act on inaccurate information, then the police risks losing the public's trust and would be unable to collect information vital to preventing crime. Mr. Iida, in the government role, agreed with Mr. Tsubohara's assessment. He noted that for AI to be widely accepted by society to advance digital transformation, it is vital to explain technology to the public in an easy-to-understand manner.

Dr. Sano, who spoke from a human rights and social justice standpoint, pushed back on students' consensus that AI should be encouraged to inform decision making in business and government settings. Noting that artificial intelligence relies on the accumulation of large amounts of data, Ms. Sano problematized Japan's relative lack of statistical data and data collection, especially around variables related to social inequalities such as gender or nationalities. Robust datasets are necessary to understand and analyze social issues, and Ms. Sano suggested that a higher priority for Japan at this current time would be to accumulate these datasets.

Ema, the moderator, noted that the exercise accounted for a limited number of stakeholders, and excluded, for instance, store employees or the retail facility owners. Ema stressed that a robust multi-stakeholder discussion should consistently ask "who is not in the room."

To conclude the panel discussion, each mentor was asked to leave a message for participants. Mentors praised participants for proactively crossing disciplinary boundaries during discussion, and suggesting concrete steps forward. Mr. Tsubohara and Mr. Iida both encouraged students to take AI ethics matters into their own hands, rather than waiting for authoritative institutions such as the government and tech companies to dictate these rules. Dr. Sano and Mr. Furukawa encouraged participants to broaden their expertise to multiple fields from an early stage in their careers. The mentors viewed the workshop as a positive opportunity for students to engage with AI ethics issues, debate solutions and changes, and drive acceptance and innovation of AI in society.



**Yoichi Iida**

Director, Information and Communication Policy Research, International Strategy Bureau, Ministry of Internal Affairs and Communications



**Atsuko Sano**

Project researcher, Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Part-time Lecturer, Graduate School of Social Design Studies, Rikkyo University

### 3.3.4 Comments from a Media Perspective (Mr. Ryohei Yasoshima, Nikkei Shimbun)

To conclude the workshop, Mr. Ryohei Yasoshima, a journalist on digital policy with the Nikkei Shimbun, spoke to participants about how he would draft an article about the fictional scenario and the ensuing dialogues if he were assigned to cover the case. He explained that the scenario would lend itself to two types of articles. First, he may initially release a short, easy-to-read “breaking news” article indicating that the government is considering the use of FRT for identifying terrorism suspects. Then, perhaps in the following weekend, a longer article may follow providing in-depth analysis and viewpoints of various people and organizations reacting to the news. In particular, Mr. Yasoshima perceived that a key issue was transparency, so he would feature concerns about privacy from individuals, as well as various organization’s calls for transparent, monitoring mechanisms.

Furthermore, Mr. Yasoshima anticipated that he would closely follow ongoing development on the story, and speak with people as diverse as salarymen on the street, tech companies, police, and government officials. Because media representation can be skewed just as datasets for algorithms can be skewed, Mr. Yasoshima emphasized the importance of stakeholders in transmitting their positions so that it can be picked up by major media outlets, and encouraged participants to transmit their ideas using social media platforms such as Note (a Japanese blogging platform).

## 3.4 Post-Workshop Evaluation Results

### 3.4.1 Overall Satisfaction with the Workshop

Thirty students, or 85% of participants, responded to the post-workshop survey. Overall, feedback for the workshop was positive, with 93.3% of participants reported being satisfied with the workshop (very satisfied, somewhat satisfied), and all participants indicating they would recommend the workshop to friends and colleagues (77.8% certainly recommend, 22.2% somewhat recommend).

The most critical feedback was on time allotment, where 40% of respondents were dissatisfied. Free responses indicated that the multi stakeholder interactive activity was too short. Additionally, the Breakout feature on Zoom caused some confusion as participants were assigned to the wrong rooms when switching through different stages of the interactive activity.

### 3.4.2 Student Learnings

Participants showed promise in becoming leaders who can bridge different fields and consider ethical perspectives to technology. When asked how students plan to be involved in AI ethics in the future, students expressed interest in conducting cross-disciplinary research in topics such as incentive design for businesses or definitions of a fair dataset. Additionally, 100% of respondents answered that they are able to consider AI issues from a number of different perspectives (47% “Definitely,” 53% “Somewhat”).

To identify how the workshop shifted students’ perceptions of AI ethics, the survey



**Ryohei Yasoshima**

Digital Policy Editor, Nikkei Shimbun Editorial Bureau

asked students to fill the sentence, "I used to think [blank], but now I think [blank]." Answers indicate that the workshop helped students understand the potential impact of having inclusive, diverse conversations around AI ethics, and built their confidence to participate even if they had little previous knowledge. Below are selected responses:

*" I used to think that AI had many issues, legal frameworks were far behind, and applications had to wait, but now I understand that my role is to participate in the public conversation and create AI-related policies & legal frameworks together. "*

*" I used to think that I only had a vague understanding of AI, but now I have more confidence in participating in debates surrounding AI and want to learn more about AI governance."*

*" I used to think that facial recognition had many benefits, but now I think that we should respect and listen to opinions of people who occupy different viewpoints and positions than I do."*

*" I used to think that the public sector and technologists should take the lead on discussing AI, but now I think that ordinary citizens and many other stakeholders must actively participate in discussion in order to create a human-centered AI-enabled society. "*

The survey had encouraging results for one of the longstanding challenges for facilitating multi-stakeholder dialogue: a gap in digital literacy between technical and non-technical students. After the workshop, all but 2 respondents responded that they would be able to explain how AI can reflect human biases, indicating that most students understood enough of the basic mechanism of artificial intelligence and its reliance on datasets. This result demonstrates the utility of lectures from the first day which was intended to provide students with a baseline of minimum knowledge necessary for a constructive discussion.

# 4 Discussion and Future Directions

---

## 4.1 Local Contexts and Global Debates: Trust, Policing, and Surveillance

This workshop asked participants to discuss uses of FRT in law enforcement and business purposes through a hypothetical scenario. This took place against an international background where, following the resurgence of the Black Lives Matter movement in 2020 following the murder of George Floyd in the United States, a number of American technology companies announced temporary or permanent moratoriums on the sale of facial recognition technology to law enforcement. Outside the United States, debates are also taking place within AI ethics scholarly communities over the use of FRT in law enforcement<sup>[15]</sup>. Many take a skeptical position on expanding its applications to the police. In Japan, there has yet to be a wide debate over FRT in the police with a similar sense of immediacy as the United States.

It goes without saying that police in Japan have a different history, level of public trust, and relationship to race than police in America. However, as mentors and participants also pointed out, an important goal of this workshop is to contribute to protecting the rights of minority groups, and work towards building trust and dialogue between government, police, and residents, in concurrence with global conversations over FRT, surveillance, and the role of public institutions.

Additionally, thanks to this workshop being organized with UNODA, this workshop was able to bring in perspectives that are often overlooked in Japan in at least two respects. First, by featuring a scenario on counter-terrorism that has not seemed as urgent in Japan as other regions, the workshop was able to introduce points of contention over race and technology that are more prevalent abroad. Should another scenario have been chosen, students would not have had the opportunity to ask and understand controversies over FRT and race in conversations outside of Japan.

Additionally, this workshop enabled a space to discuss gender and technology. As Dr. Sano noted, many feminist organizations in Japan have not yet incorporated algorithmic justice into their agenda. The workshop gathered participants who had an interest in gender, including leaders of feminist organizations in Japan, and suggested that AI could be one lens for understanding the impacts of gender and methods of inclusion.

On the other hand, concerns related to “access to technology” did not feature in participant discussions on the hypothetical scenario, although the theme is often brought up in international discussions. This may reflect a collective blind spot about digital divides, or reflecting a belief that Japan is a country with a small digital divide.

[15] For example, at the ACM-FAT conference held in Barcelona in 2020, there was a debate over the tutorial “Can an algorithmic system be a ‘friend’ to a police officer’s discretion?”

## 4.2 Notes on the Framing of the Discussion

While critical viewpoints are being exchanged internationally regarding use of FRT in police, this particular workshop chose to ask students to reach a “consensus” among stakeholders on whether the police should utilize FRT.

At first glance, the idea of “consensus” seems to risk excluding minority opinions. However, prompting participants to reach “consensus” could shift the focus of the workshop to propose what constraints and conditions would be necessary for AI use. As FRT is being moved into implementation in various settings, there are trade-offs to be made in relation to privacy, security, among other objectives. Rather than to present binary choices between A and B, the organizers aimed to suggest a more flexible way of thinking where certain conditions and constraints would make either A or B possible. Thus we presented a discussion framework that asked participants to reach consensus on either of the extreme choices between “Yes” and “No” to the use of FRT.

Certainly there are some technologies that warrant debate over whether they should be built at all, including lethal autonomous weapons systems (LAWS) that the UN has problematized. Accordingly, it will be important going forward to discuss the appropriateness of the questions being used to frame the discussion, along with the improvement of the hypothetical scenario and overall program.

## 4.3 Future Directions

A few issues must be addressed for similar workshops in the future. As this was an online workshop, there were issues with time allocation and smooth configuration of Zoom Breakout rooms. Additionally, mentors had to be relied upon for heavy support, because content from Day 1 and Day 2 did not necessarily build on one another in a clear manner, and ambiguous details from the hypothetical scenario invited confusion (for instance, whether the counter-terrorism system is built from scratch or is extended from pre-existing commercial systems).

As for future workshops, there may be three potential future directions. First, due to the online format of the workshop, there were participants from outside of Tokyo, and even from abroad. Thus it may be possible to extend this workshop format and hypothetical scenario to open discussion in AI ethics to an even wider audience. In particular, given the positive impact the workshop had on participants’ willingness to engage in AI governance, it may be effective to target populations who have been historically excluded from AI development, such as the elderly and people with disabilities.

Furthermore, there is a possibility of expanding this workshop overseas. Past UNODA workshops for STEM students have not included interactive simulations of multi-stakeholder discussions. Since discussions about AI vary depending on local context, the workshop format may be adapted to other settings. Finally, it may be possible for this workshop to be directed towards companies and public agencies. According to a survey of companies there is a possibility for this

workshop to be used as AI governance training for companies and governments. According to a survey of companies who have provided AI surveys domestically and abroad, those who have responded they have succeeded in AI applications were found to have conducted AI Ethics training in-house<sup>[16]</sup>. As issues such as fairness in AI grow in urgency, it would be important to prepare a variety of educational resources.

While such developments are possible, discussion on AI fairness and governance is not only about its content, but also about input on various social issues, and issues remain to be solved to make it a general-purpose educational program. The organizers will continue to improve upon the program through trial and error, and hope that this report will be helpful to those who are considering education programs on AI fairness and governance.

[16] <https://news.mynavi.jp/article/20181024-712503/>

## Authors' Profiles



**Arisa Ema** is a Project Assistant Professor at the Institute for Future Initiatives at the University of Tokyo. She serves as a member of the Ethics Committee in the Japanese Society for Artificial Intelligence

---



**Moe Sunami** is a member of the Class of 2022 at Pomona College in the United States, where she majors in mathematics and philosophy. She is supported by scholarships from the Yanai Tadashi Foundation and Son Masayoshi Foundation, and is a former app developer.

---



**Shiori Komatsu** graduated from University of Tokyo, Faculty of Law in 2020. She passed the Japanese Preliminary Bar Examination in 2019, and the Bar Examination in 2020. She is a member of the World Economic Forum Global Shapers Community, serves as a Japan-China Friendship Youth Ambassador, and is a researcher with the AI and Law Society Research Group. She is a member of the Son Masayoshi Foundation.

---



**Charles Ovink** is a Political Affairs Officer with the United Nations Office for Disarmament Affairs (UNODA) and its Regional Centre for Peace and Disarmament in Asia and the Pacific (UNRCPD). He is also a program manager at the United Nations University World Institute for Development Economics Research (UNU-WIDER).

---



**Takashi Matsumoto** is a visiting researcher at the Institute for Future Initiatives at the University of Tokyo, and a manager at Deloitte Touche Tohmatsu LLC.

---

## Appendix 1: Workshop Scenario

[1] A Japanese IT company “X” has developed in-store cameras equipped with a facial recognition and a behavior detection system, which can notify store employees in real time of customers who display suspicious behaviors suggestive of shoplifting. The system is to be installed in large, private retail facilities across Japan.

[2] Company X has explained that its technology has been developed with guidance from crime prevention experts and behavioral data from the field. The content of the algorithm has not been disclosed, so it is not known which features are used for identifying suspicious customers (i.e., their face is hidden, or they are carrying a large package).

[3] Large private retail facilities with these cameras installed have praised the technology, saying that it enables employees to monitor flagged customers more closely. This has reduced the economic cost of shoplifted items, and decreased the need for overtime work.

[4] Additionally, these retail facilities have requested a feature that can track and monitor the behavior and number of visits for customers who have registered their facial data with the system. Registered customers would also be assigned redeemable “Eco Points” if they display environmentally conscious behavior, such as bringing reusable bags or cups, and purchasing environmentally friendly products. This feature is anticipated to benefit both the store and the customer by encouraging both purchases and engagement with the environment.

[5] During the same time period, a number of bombings have taken place in Japan by unknown suspects. These attacks have occurred in large, retail facilities and subway stations frequented by many people.

[6] In response, the government has initiated a drastic counter-terrorism initiative, and requested that Company X’s technology, which is already in use across the country, add features that can extract behavior that is suggestive of terrorism and identify suspects.

[7] Company X has said that this feature would be built based on actual field data. Since the content of the algorithm has not been disclosed, it is unclear what features are used to identify terrorism suspects (e.g., covering the face, carrying a large package). Law enforcement will not provide information on specific, wanted terrorism suspects to Company X.

[8] The government is considering requiring large retail facilities to install Company X’s systems, and report its results to the police in real time.

[9] Additionally, the government is also considering granting law enforcement officers with the authority to conduct searches of personal belongings on individuals that the AI system identifies as suspicious.



## Discussion Questions

For each of the three questions below,

1) During discussion within your assigned stakeholder group:

Organize arguments from your own stakeholder's perspective.

2) During discussion with other stakeholders:

After explaining your own stakeholder's position, aim to reach a consensus (and the conditions for consensus) among the multiple stakeholders present.

Q1: Company X's facial recognition system should continue to be used commercially in the future.

Yes / No

Q2: A counter-terrorism feature should be added to Company X's facial recognition system.

Yes / No

Q3: As part of digital transformation (DX), the public and private sectors should be encouraged to use AI to inform their decision-making.

Yes / No

## Presentation Guidelines

Each group will have 3 minutes to present. Please edit and screen share the following slide template.

<b>Multi-Stakeholder Group (1)</b> <b>❶ Company X's facial recognition system should continue to be used commercially in the future.</b>  Yes / No	<b>❷ A counter-terrorism feature should be added to Company X's facial recognition system.</b>  Yes / No	<b>❸ As part of digital transformation (DX), the public and private sectors should be encouraged to use AI to inform their decision-making.</b>  Yes / No
Conditions for consensus (binding laws, non-binding guidelines, etc.)	Conditions for consensus (binding laws, non-binding guidelines, etc.)	Conditions for consensus (binding laws, non-binding guidelines, etc.)

## Key Stakeholder Values

### Government

Public safety, Convenience for residents, Strengthening of the country through new technologies, Coordination with international trends

### IT Company "X" Providing Facial Recognition Technology

Safety, Explainability, Accountability, Market share, Sustainable Operations

### Civil Society (Human Rights)

Discrimination and inequality, Privacy, Surveillance-free society, Transparency, Digital divide

### Police

Protecting the lives and physical safety of residents, Productive use of police power

## Discussion Guidelines

### Overall

- Keep your opinions concise, and try to give as many people a chance to speak as possible
- Instead of arguing with others as adversaries, work together to find good arguments
- Feel free to pose different opinions from others, and come up with a variety of arguments
- Be respectful of other participants, refrain from making personal attacks

### Discussion within Stakeholder groups

- Begin by thinking about what your stakeholders value
- No matter what your opinion is, work from the perspective of trying to understand the values of your own stakeholder group
- Instead of looking to the mentor to give answers, think together with other participants

### Discussion within Multi-stakeholder groups

- Combine the ideas you had discussed within your own stakeholder group with those of other stakeholders, and find better arguments
- If you change your mind during discussion, it is a sign that you gained another perspective and are thinking carefully
- Affirm the value of each stakeholder, listen to different opinions, and cooperate with each other to find good arguments



**United Nations Office  
for Disarmament Affairs**



**東京大学未来ビジョン研究センター**  
Institute for Future Initiatives, The University of Tokyo

