

**Study Group 'AI governance and its Evaluation'  
Report on the Session #2**

**1. Introduction**

The Japan Deep Learning Association establishes study groups as a forum for deepening knowledge and discussing domestic and international policy trends related to artificial intelligence (hereafter AI) and Deep Learning (hereafter DL). This study group, "AI Governance and its Evaluation," defines "governance" as a system of management and evaluation by various actors, and launched a study group in July 2020 to investigate what forms of governance are possible and conduct a year-long study to help build trustworthy AI systems.

In the second session, the vice chair of the study group Mr. Takashi Matsumoto (Deloitte Touche Tohmatsu LLC / Institute for Future Initiatives, the University of Tokyo) gave a presentation on the "AI Ethics Guidelines" overview.

This report is a reconstruction of the topical presentation and the discussions of the study group participants.

**2. "AI Ethics Guidelines" for AI Governance**

**Background & main issues of AI Ethics Guidelines**

In recent years, while the implementation of AI in society has been accelerating, ethical issues have arisen. Incidents such as the wrong transmission of domestic conversations by AI speakers and the alleged female discrimination by AI in personnel recruitment have made the ethical reliability of AI itself, in terms of fairness, privacy, and transparency, an important issue.

Given the above background, policies, guidelines, and other reports related to AI ethics have been presented, and discussions and research and development have been actively conducted internationally. At the same time, it has become clear that there is a gap in interpretation of the elements of AI ethics due to differences in country-level policies, global issues and practical perspectives.

Moreover, there are cases where the elements pertaining to AI ethics defined in each AI ethics guideline have different meanings and interpretations even when they are expressed in the same way, making common recognition difficult (See the table below).

■ Consideration in interpreting the AI ethics guidelines (Excerpted from the public materials of this study group)

Example: "Transparency"

Name of guideline	Interpretation of "Transparency" in the guideline
"AI Utilization Guidelines" in Japan	Including "Traceability" and "Explainability"
"Ethics Guideline for Trustworthy AI" in EU <sup>1</sup>	Including "Traceability" and "Explainability" plus, "Communication"
"Model AI Governance Framework" in Singapore <sup>2</sup>	Treating "Explainability" and "Transparency" as separate elements.

Furthermore, as the elements of AI ethics such as "Accountability", "Robustness", "Fairness", "Transparency", are not completely independent, but have an interrelationship, considerations for organizing them structurally have been made. For example, in a "Survey on Trends in AI Governance<sup>3</sup>" conducted by the Ministry of Economy, Trade and Industry (METI)<sup>3</sup>, the elements are categorized into three areas: "ethical elements required as the basis" for service provision, "ethical elements related to AI decisions" for data used by AI and AI models, and "ethical elements related to the users" for AI output. It is mentioned that each area is interrelated.

### Overview of AI ethics guidelines & case studies

Overseas, international organizations, OECD<sup>4</sup> and UNESCO<sup>5</sup> have both adopted "AI Principles," the European Commission<sup>6</sup> has proposed "Ethical Guidelines for Trustworthy AI," and Personal Data Protection Commission Singapore has presented a self-assessment guide on the risks posed by AI<sup>7</sup>. In Japan, as represented by the Cabinet Office's "Social Principles of Human-centric AI" and Ministry of Internal Affairs and Communications' "AI Utilization Guidelines," industry, government, and academia have released a number of guidelines on AI ethics. In the private sector, foreign companies such as Microsoft<sup>8</sup>, Google<sup>9</sup>, and IBM<sup>10</sup>, as well as international Japanese companies such as SONY<sup>11</sup>, Fujitsu<sup>12</sup>, and NEC<sup>13</sup> have released guidelines on AI ethics.

<sup>1</sup> <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

<sup>2</sup> <https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework>

<sup>3</sup> <https://www.meti.go.jp/metilib/report/2019FY/000199.pdf>

<sup>4</sup> <https://oecd.ai/ai-principles>

<sup>5</sup> <https://en.unesco.org/artificial-intelligence/ethics>

<sup>6</sup> <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

<sup>7</sup> <https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework>

<sup>8</sup> <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3apriorityr6>

<sup>9</sup> <https://ai.google/responsibilities/>

<sup>10</sup> <https://www.ibm.com/blogs/think/jp-ja/everyday-ethics-for-artificial-intelligence/>

<sup>11</sup> [https://www.sony.net/SonyInfo/csr\\_report/humanrights/AI\\_Engagement\\_within\\_Sony\\_Group.pdf](https://www.sony.net/SonyInfo/csr_report/humanrights/AI_Engagement_within_Sony_Group.pdf)

<sup>12</sup> [https://www.fujitsu.com/global/documents/about/csr/humanrights/fujitsu-group-ai-commitment-201903\\_en.pdf](https://www.fujitsu.com/global/documents/about/csr/humanrights/fujitsu-group-ai-commitment-201903_en.pdf)

<sup>13</sup> <https://www.nec.com/en/press/201904/images/0201-01-01.pdf>

In addition, standardization organizations such as IEEE and consortiums such as "Partnership on AI<sup>14</sup>" are actively engaged in research and development as well.

The AI ethics guidelines issued by each organization are formulated according to global issues, policies and practical perspectives of each country, and contain recommendations and considerations not only for AI developers and providers, but also for users. However, consideration from the perspective of users of AI in business is not sufficient enough yet, and discussions including user companies are needed in order to ensure the reliability of ethics in the entire supply chain of AI services.

■ Characteristics of AI ethics guidelines by issue sources (excerpted from the public materials of this study group)

Issued by	Characteristics of the AI ethics guidelines
International Organizations	Recommendations to governments → Focus on social issues such as participation of least developed countries, digital disparity, and environmental issues
Countries & Governments	Formulated in line with ethics guidelines of international organizations → Contain recommendations for structured approaches (frameworks, assessments) and collection of corporate case studies
Companies	AI development companies formulate "action guidelines" plus, technical implementation methods are also disclosed overseas → Consideration on the part of <u>AI users in business</u> is insufficient
Organizations & Community	Standardization and research focused on specific areas (transparency, fairness, human interaction, etc.)

### Towards Principle to Practice

In applying the AI ethics guidelines issued by various organizations to actual AI services, the following points can be cited as influences on the social activities of each company.

The first point is the impact on ethical activities in companies. In recent years, many companies have been promoting compliance activities to ensure ethical behavior, led by ethics officers and leaders of each business unit. Mr. Matsumoto mentioned that the need for "sharing values," "governance," and "penetration of AI ethics in the organization" as roles required of leaders who promote such ethical activities.

The second point was from the discussion at the expert meeting held as part of the Ministry of Economy, Trade and Industry (METI)'s survey on trends in AI governance<sup>15</sup>.

<sup>14</sup> <https://www.partnershiponai.org/human-ai-collaboration-framework-case-studies/>

<sup>15</sup> <https://www.meti.go.jp/metilib/report/2019FY/000199.pdf>

As an impact on the development lifecycle of AI services, the necessity of AI governance at the time of "before AI service release," "after AI service release," and "response to incidents" was cited.

The third point raised was the impact on the relationship between companies and each stakeholder after the formulation of AI ethics guidelines (see the table below).

■ Impact on the relationship between companies and each stakeholder (excerpted from the public materials of this study group)

Stakeholders of the company	Impacts with the stakeholders listed on the left
Industry associations/ Regulators	Compliance with industry ethics guidelines
Business partners/ Contractors	Understanding of corporate ethics guidelines
Users/ Customers	Understanding the purpose of use and user responsibilities
Third party/ Society	Understanding and acceptability of social impact

### 3. Main comments from the workshop participants

As mentioned above, the second session addressed the "Ethics Guidelines" for AI governance. Based on the AI ethics guidelines issued by various organizations, an overview of each guideline and case studies were presented, and the impact on each company's activities in the stage of applying them to actual services was discussed.

The following comments were made by the participants based on the topic presented.

- ✓ Since most business-to-business (B2B) transactions in Japan have a so-called B2B2C (Business-to-Business-to-Consumer) structure, and the parties involved in AI services are often made up of multiple business entities, a framework for dialogue among the parties involved is necessary for AI service providers to comprehensively consider multiple risk factors.
  
- ✓ In the U.S., there have been multiple problems caused by AI services, and companies, especially in Silicon Valley, are taking a risk-based approach. Provided that the primary objective is to maximize corporate value, and this is not a one-size-fits-all approach, as different industries have different approaches. Since there is also the issue of cost-effectiveness, it was pointed out that it is similar to the ESG investment story in that it can be incorporated if it is deemed necessary to increase corporate value (for example, if it is an EU partner, it will emphasize privacy protection). Therefore, as with security and anti-harassment measures, with regard to ethics, there is a tendency to emphasize growth while ensuring a minimum level of compliance, and to

address risks along with scaling up. However, it was also pointed out that the awareness of issues related to racial and gender discrimination and transparency is culturally fostered in Europe and the United States, and that the prerequisite stage may be different from that in Japan.

- ✓ The Ministry of Internal Affairs and Communications (MIC)'s "AI Utilization Guidelines" and other guidelines do not only target "ethics," but also comprehensively include issues to ensure the reliability of AI services, such as "safety," in order to promote the utilization of AI and ultimately increase corporate value. Thus, a comprehensive response to various risks, not limited to "ethics," is required in actual services.
- ✓ In business, it is necessary to be compatible with corporate governance. Consistent consideration is needed from management's vision to field operations and communication with each stakeholder.
- ✓ Not just overseas, but some Japanese companies are starting to provide in-house training on AI governance and data governance. Their materials are created in the form of security and harassment training. In addition to employee training, explanations to customers were also pointed out as an issue.

We will continue to discuss AI governance in Japan and abroad through this study group.

Written by Keitaro Saito

Translated by Michiko Shimizu

<Outline of the 2<sup>nd</sup> Session of the Study Group>

Date & Time: Tuesday, August 25, 2020, 16:00-18:00 (Zoom)

Agenda:

- Topic: "AI Ethics Guidelines Overview" provided by Takashi Matsumoto (Deloitte Touche Tohmatsu LLC/ Vice chair of the study group)
- Question and answer session / discussion