June, 2021

JDLA

Study Group 'AI governance and its Evaluation'

Report on the Session #14

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 14th session (May 11, 2021), Mr. Kodo Shu of Huawei Technologies Japan K.K. and Mr. Naohiro Furukawa of ABEJA Inc. presented topics on the theme of practices on AI governance in businesses.

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

2. Practice on AI Governance at Huawei

Mr. Shu presented a topic titled "Responsible AI: Huawei's Recommendations on AI Governance".

Huawei's holistic view of AI governance

Huawei believes that the main point of AI governance should not be to overly restrict innovation, but to promote exploration, reliability, and control over new technologies. Therefore, it is necessary to define various roles in AI systems for each actor in AI governance, and to develop action plans according to the roles, so that actors understand and recognize their own responsibilities (obligation).

Huawei's proposed actions on AI governance

In building a trustworthy AI, Huawei proposes three actions as shown in Table 1 below.

Table 1: Huawei's proposed actions on AI governance[1]

Action Proposal 1	➢ Build an international AI governance platform, establish a professional permanent or non-permanent international governance organization which incorporates public and private members from each region to realize a public-private joint effort to promote cooperative behavior among stakeholders and advance technology & industry based on fair order.
Action Proposal 2	➢ From a hierarchical perspective, provide targeted governance guidelines and best practices for actors at different tiers of the AI system. ➢ Collect and organize AI governance "best practices" across tiers and provide minimum acceptable standards for governance at each tiers.
Action Proposal 3	➢ For high-risk scenarios and applications, acquire the necessary certification by an independent third-party in accordance with the principle of necessity and minimization. ➢ For low-risk scenarios and applications, apply voluntary certification. ➢ Apply the "principle of limited liability" for the approved AI systems. ➢ Refer to the existing conventional ICT product safety and reliability standard system and certification mechanism, and directly divert a portion if possible. ➢ Establish information disclosure mechanisms and evaluation certification criteria for AI governance, and define standards and scope for information disclosure.

The background to Action Proposal 1 is that various organizations around the world, including private companies, industry associations, and international organizations, have issued "AI Ethics Guidelines and Principles." While respect for diverse values is a necessary prerequisite, an overly fragmented governance framework may hinder the development of innovation and even lead to Regulatory Arbitrage[2]. In response to the circumstances, it is required that governments, civil society organizations, and actors involved in AI governance to work together to build an international AI governance

[1] Excerpts from the public materials of this study group.

[2] Regulatory arbitrage is a practice of circumventing unfavorable regulations by moving from one heavily regulated jurisdiction to another with more lenient rules.

system (see Table 2), to form a basic consensus, and to coordinate and collaborate on areas of pluralism and differences.

Table 2: Characteristics and functions of international AI governance system[3]

Recognition and support by government agencies	Government agencies in each region establish the developed AI ethics principles and governance framework as a regional one.
Promotion of public-private joint effort	Promote "public-private joint governance (public-private joint effort)" by utilizing the strengths and know-how of the private sector in technology R&D and establishment of standards.
Promotion of international standardization	Establish international standards and norms (especially minimum acceptable baselines) for safety, reliability, etc. in the field of AI technology in collaboration with standardization bodies.
Establish an authoritative certification testing mechanism	Authorize independent international testing organizations to conduct evaluation tests of AI products/services based on uniform test specifications to ensure that the products/services meet market entry requirements.
Promotion of practical application of technologies	Promote the use and development of AI in various industries by collecting information on market needs and opinions.

3GPP[4] is a successful example of an international governance mechanism. 3GPP's membership includes a number of MRP[5]s, which are responsible for facilitating consensus building on 3GPP's mobile communication technologies by providing technical requirements and market opinions, and promoting the use and development of global mobile communication networks. 3GPP works with the OP[6]s to determine the overall policy and strategy of 3GPP, and develops common technical specifications on behalf of the OPs. As a result, each OP needs to refer to 3GPP-related standard specifications when developing its own standards.

---

[3] Excerpts from the public materials of this study group.

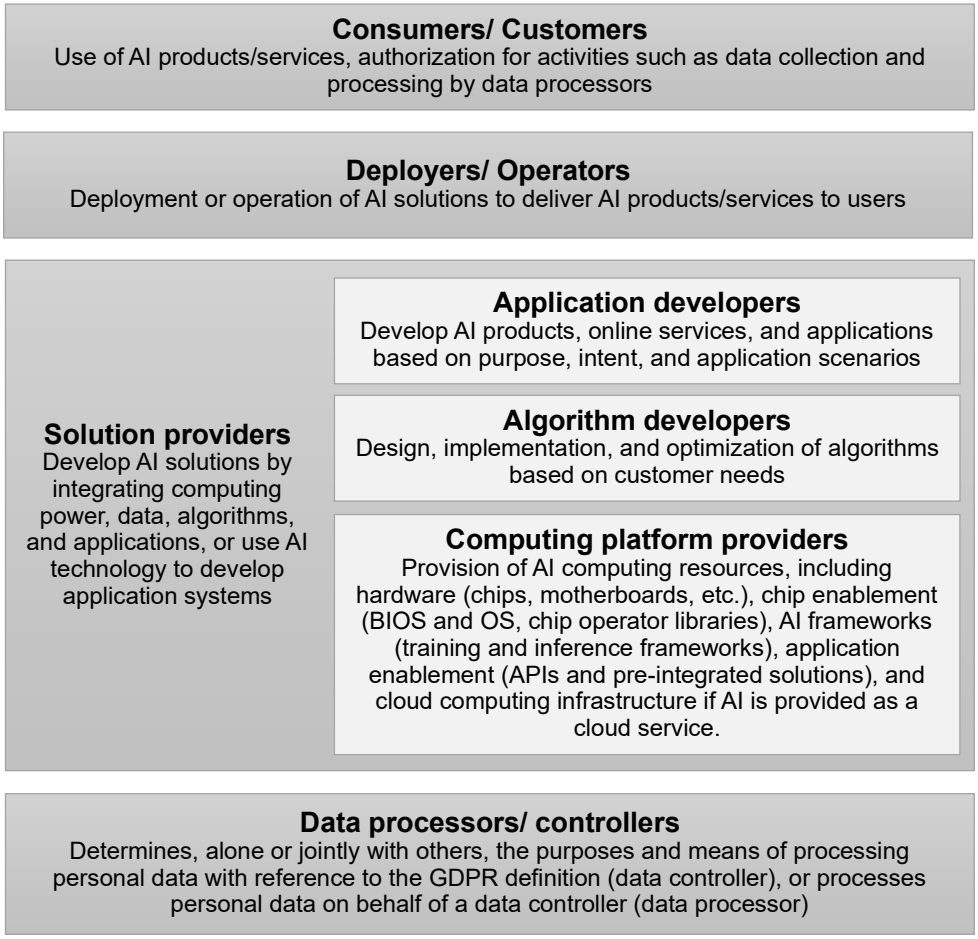
[4] 3GPP, 3rd Generation Partnership Project is a collaborative project launched in December 1998 with the initial goal of developing globally applicable specifications for third-generation (3G) mobiles systems. 3GPP is composed of national and regional standardization bodies and is currently responsible for promoting research and standardization of mobile communication technologies.

[5] MRP stands for Market Representation Partners, and in this case, market representation partners for mobile communication technology.

[6] OP stands for Organization Partner, which in this case is a standardization organization partner for mobile communication technology.

The key point in Action Proposal 2 is the establishment of a hierarchical governance framework for AI systems (hereafter referred to as hierarchical governance, see Table 3). Hierarchical governance further subdivides and stratifies the governance architecture of AI systems from a hierarchical perspective, helping to identify risks for actors at each tier (see Table 4) and take preventive measures. Hierarchical governance also enables collaboration with standardization bodies to develop global standards and norms for best practices in AI governance at different tiers. It also enables the building of AI testing and certification capabilities that are based on different tiers of AI systems. Therefore, it is believed that hierarchical governance will facilitate traceability across tiers of AI systems and help determine responsibilities at each tier.

Table 3: Hierarchical governance[7]



[7] Excerpts from the public materials of this study group.

Table 4: Overview of the roles and activities of actors in hierarchical governance[8]

Actors	Roles and Activities
Consumers/Customers	Have the right to choose whether or not to use the AI system; use safely by following product/service instructions and avoid using it for purposes that violate laws or ethics, such as misusing AI technology to create fake videos, fake sounds, fake photos, etc.
Deployers/Operators	Inspect AI systems for conformance to intent, identify and correct adverse outcomes, ensure implementation objectives are met, and effectively manage and prevent AI security and privacy risks during implementation and operations.
Solution Providers	Assure the realization of AI system business functions and reliability goals, and provide solutions and ancillary services that meet the safety and ethics requirements of the scenario through safeguards such as proactive protection, intermediate intervention, and post-audit evaluation.
AI Application Developers	Provide services, applications, subsystems, and ancillary operation and maintenance mechanisms that meet the reliability requirements of each sector by selecting AI algorithms and using computing platforms such as cloud, edge, and devices that meet security standards.
AI Algorithm Developers	Develop models and algorithms that meet security and robustness criteria, and continuously improve the ability of algorithmic programs to perform as expected by using methods such as statistical analysis and logic verification.
AI computing platform provider	Build a reliable computing platform with information security enhancement hardware, security chips and defense components for machine learning, reliable operator realization, AI operational framework and AI application enablement, etc., to provide a safe and robust operating environment, and guarantee traceability, privacy protection, safety, and robustness of data and running programs.
Data processors/ controllers	Ensure that data management and operations are fully compliant with applicable regulatory requirements, such as GDPR.

[8] Excerpts from the public materials of this study group.

The key point in Action Proposal 3 is to introduce certification requirements according to risk scenarios and AI applications. The EU's February 2020 “White Paper on Artificaial Intelligence[9]” specifies criteria for identifying high-risk scenarios and AI applications,[10] but the introduction of mandatory certification schemes could unduly increase compliance costs and stifle innovation and technology development. In light of this, following the principle of minimization would be a more effective way to set up high-risk scenarios and AI applications target list. Having said that, there are actually not many cases where AI applications have high-risk scenarios in practice. For low-risk scenarios and AI applications, the system encourages businesses to improve their quality and sophistication by optimizing the voluntary certification program and introducing the "self-declaration" method through classification management based on the company's management level and credibility.

Further, some of the existing ICT product safety and reliability standard systems and certification mechanisms can be directly appropriated in building the testing and certification capability for AI applications. Conventional ICT product system engineering has been built on a step-by-step testing and certification system, and the process is very mature. In addition, the ICT industry has many specialized international certification bodies in addition to standardization bodies, and these certification bodies can objectively evaluate the security and privacy protection of ICT products.

3. Practice on AI Governance (ABEJA)

Next, Mr. Furukawa presented a topic titled "Practice on AI Governance in ABEJA".

ABEJA's Governance Structure

- **Personnel and Organization for AI Ethics**

  ABEJA assigns AI ethics practical tasks to its legal staff. Due to the importance of AI ethics and management’s awareness of it[11], in 2019, ABEJA inaugurated a committee, “Ethical Approach to AI (EAA)”[12] where external intellectuals discuss issues related to AI in ethical and legal perspectives.

- **AI Ethical Policy**

  ABEJA has not established a specific policy on AI ethics. The main reasons for this are the following two points.

---

[9] https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[10] Examples of high-risk scenarios include healthcare, transport, and energy. Also, the use of AI applications in recruitment and AI applications used for remote biometric identification are always considered "high-risk".

[11] https://abejainc.com/ja/news/article/20190821-2542

[12] https://abejainc.com/ja/news/article/20190725-2522 (in Japanese)
https://abejainc.com/en/news/article/20190725-411 (in English)

1) While it is possible to establish ABEJA's own policies for convenience by referring to the various policies on AI ethics that are currently publicly available, ABEJA believes that concrete actions are more important than the policies themselves in bringing about ethical improvements within the company.
2) Management's awareness of AI ethics is well known within the company, and the importance of AI ethics is also shared among employees. Data scientists and project managers (PMs) in the company have a strong interest in AI ethics, and it seems that the awareness of the importance of AI ethics has taken root among employees.

- **Process for resolving issues related to AI ethics**

At ABEJA, the legal staff identifies issues related to AI ethics at the time of initial project contract review and before NDA[13] signing.

The process of identifying, examining, and solving issues is as follows.

1) The legal staff identifies the presence or absence of issues related to AI ethics from the project proposals, and interviews the project person in charge case as necessary.
2) When issues related to AI ethics are identified, the legal staff will have a meeting with the PM in charge, and the PM will discuss with the client the points to keep in mind regarding the issues.
3) If issues are identified that may require the views of external experts, they will be treated as discussion topics for the EAA.

ABEJA approaches problem solving from the stance of whether or not it can provide services that customers and their end users require.

- **Information sharing and training on AI ethics for employees**

At ABEJA, news related to AI ethics is routinely shared and discussed on Slack[14], and internal study sessions are held several times a year to share and present the contents of studies at EAA and social trends in AI ethics.

Discussion topics in EAA

Among all the areas of compliance in the company, EAA talks over issues on "AI-related laws and ethics". In order to secure objectivity and independency, the committee is formed by external intellectuals (hereinafter referred to as EAA members) entirely. From

[13] NDA stands for Non-Disclosure Agreement
[14] Slack is a business chat tool developed and provided by Slack Technologies, Inc.

ABEJA, President & Representative Director, CEO, the company secretariat, and those in charge of project matters depending on the discussion theme participate in the EAA discussions.

The main discussion topics of the EAA are as follows.

1) Advice and proposals for each issue cases

   In the committee, members discuss open issues related to AI and ethics that ABEJA comes across in business operations. The committee gives constructive advice and suggestions that may lead to enhance and improve ABEJA's operation.

2) Make internal principles and guidelines on AI usage

   The committee aims to make AI usage principles and guidelines to be shared within ABEJA. It will reflect the trend of AI ethics guidelines, etc., in Japan, EU, etc., and opinions by ABEJA members. There will be some opportunities where ABEJA members and the committee exchange their opinions prior to the establishment of the principles and guidelines.

3) Share knowledges and insights among intellectuals

   When the committee works on outlining guidelines and having discussions on issues, the committee members share their viewpoints in terms of overseas politics/law trends and their expert insights. It is to vitalize discussion and to be referred to when they make judgements.

When submitting AI ethics issues related to ABEJA's individual projects for discussion at the EAA, the NDAs with customers will be taken into account and generalize the discussion topics.

Awareness of issues in the development of AI ethics policies

Based on the current status of various publicly available policies on AI ethics, it seems that AI governance in B2B (Business-to-Business) companies, start-ups, and AI vendors is uncertain, and that this situation is not accidental.

- **Awareness of the problem from the perspective of a B2B company**

  It seems that the content of the guidelines for AI ethics in the U.S. and Europe reflects the premise of B2C (Business-to-Consumer) companies. This is probably due to the fact that there are many companies in the U.S. and Europe that mainly develop their own software, and therefore there is no international research on practices in B2B companies. In addition, it is difficult for B2B companies to talk about AI ethics unless the customers they provide services to are aware of the ethical issues. Therefore, it is necessary for B2B companies to take actions such as raising the issue of AI ethics from them to their customers and society, and sharing and disclosing information

about their AI governance initiatives.

- **Awareness of the issues from the perspective of a start-up company**

  The current policies on AI ethics in the public domain do not seem to be designed for start-ups. While start-ups are small and have limited resources in terms of personnel and funds, they are able to share implicit values and respond flexibly and quickly. Therefore, it is necessary to raise awareness of the importance of AI ethics through the creation of an AI ethics department, the assignment of a person in charge, and internal education.

- **Awareness of the issues from the perspective of AI vendors**

  As mentioned above, in the U.S. and Europe, many companies develop their own AI products/services, and it is GAFA[15] and other in-house developers (not AI vendors) that play a central role in the AI domain. Since AI vendors are involved in a wide variety of AI development, and since the stakeholders and industry-specific cultures vary greatly from project to project (and customer to customer), there are a wide variety of ethical values to be considered. In such circumstances, while practical knowledge is accumulated, it is very difficult to build practices that abstract the knowledge.

4. Discussion points in the question and answer session

In the 14th session, practices on AI governance in businesses were discussed. Based on the topics discussed, the following questions and answers were raised.

Issues related to AI governance that need to be addressed as a top priority

- ✓ Huawei is most aware of the compliance costs incurred in dealing with the regulatory and certification systems of each countries. Therefore, it is desirable to have international standardization and certification systems in place to reduce AI development costs. In addition, the company wants to focus on business (AI development), so it is desirable to establish a system to ensure transparency in AI governance and a certification system.
- ✓ ABEJA does not have a particular preference for addressing issues related to AI ethics. Nevertheless, ABEJA's customers tend to be more concerned about privacy (how data is collected and handled), and therefore, fairness and security are not often discussed, depending on the services provided by AI.

The actors responsible for risk assessment

- ✓ While it is desirable for both the “deployers/operators” and the “solution providers”

[15] GAFA is an acronym for Google, Apple, Facebook, and Amazon, the four most powerful tech companies in the U.S.

to conduct risk assessment, as suggested by Huawei's hierarchical governance, the “deployers/operators,” who is directly interfaces with the “consumers/customers,” should proactively identify and assess significant risks. For this purpose, “deployers/operators” have an obligation to present to “solution providers” the technical specification requirements and security requirements that should be satisfied when introducing and operating AI.

- ✓ Huawei is a B2B company, which in a broad sense means that is a "solution provider”. If the “deployers/operators” have a high level of risk assessment capability, it is desirable for them to take the initiative in conducting risk assessment. If the “deployers/operators” have low or no risk assessment capability, it is necessary for the “solution providers” to clarify the scope of risk assessment through a contract with the “deployers/operators,” and then support risk assessment together with them.
- ✓ Potential risks are often raised by ABEJA, and the items to be discussed with customers are not only limited risks related to AI ethics, but also business risks. However, the final decision on the specific risk investigation and assessment is made by the customers.

The extent of information disclosure of AI technical specifications from a security perspective

- ✓ Disclosure of all AI technical specifications is not common because it increases the possibility of exposure to security risks. However, in the event of an accident, disclosure of some technical specifications will be necessary for the third-party committee that will be set up to investigate the cause of the accident.

Liability for Non-conformity (previously Defects Liability) of AI services

- ✓ Since most AI development is done under quasi-delegated contracts, there is no liability for contractual non-conformity in such cases. However, when it is found that there are differences in the AI developed and delivered by AI vendors and other businesses from the contract contents, the businesses that developed and delivered the AI need to take some action.

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 14th Session of the Study Group>

Date & Time: Tuesday, May 11, 2021, 17:00-19:00 (Zoom)

Agenda:

- Topical presentations:
  - "Responsible AI: Huawei's Recommendations on AI Governance" provided by Mr. Kodo Shu (Huawei Technologies Japan K.K.)
  - "Practice on AI Governance in ABEJA" provided by Mr. Naohiro Furukawa (ABEJA Inc.)
- Question and answer session / discussion