April 2023

JDLA

Study Group 'AI governance and its Evaluation'

Report on the Session #6 (Phase III)

1. Introduction

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 to help build trustworthy AI systems, and the Phase III began in 2022.

In the sixth session of Phase III (January 24, 2023), Hiromu Kitamura (CDLE AI legal group (NEC)) presented a topic on "AI System Auditing" for the realization of AI governance, in particular the trends surrounding regulations and standardization in Europe and the United States.

2. "Recent trends and issues related to AI System Audits" by Hiromu Kitamura (CDLE AI legal group (NEC))

Latest Trends Related to AI System Audits

It is necessary to watch trends in regulation and standardization in Europe and the U.S. as they affect future AI system audits and auditing standards. The following are some of the specific trends to be focused on.

- International standardization: The development of the AI Management System Standard ISO/IEC42001, which will serve as the basis for AI audits, is urgently needed. ISO/IEC 42006 (Information technology - Artificial intelligence - Requirements for bodies providing audit and certification of artificial intelligence management systems), which specifies audits themselves, is also being developed. In addition, the report "AI Watch: Artificial Intelligence Standardization Landscape Update (2023)[1]" issued on January 9, 2023, by the Joint Research Centre (JRC) of the European Commission, an European advisory body, specifically identifies elements in IEEE (Institute of Electrical and Electronics Engineers) standards that will meet future standardization needs. In particular, IEEE P7001 (currently named IEEE 7001 2021) is a relevant source of information to watch for in the future.

[1] https://publications.jrc.ec.europa.eu/repository/handle/JRC131155

- U.S.-EU Trade and Technology Council (TTC): Joint European and U.S. efforts are also underway: the TTC has presented a joint roadmap on evaluation and measurement tools for trustworthy AI and risk management in December 2022, and has laid out steps for U.S-EU. cooperation[2].

- EU proposal for AI Act (although relevant decisions and conclusions are unknown yet): In the European Commission's discussion of AI Act amendments as of January, it has been proposed and discussed that human oversight will be required for users of high-risk AI systems, and that an impact assessment of fundamental rights will be required to ensure compliance with AI laws. It should be also noted that the proposed amendment to the European Product Liability Directive (PLD), which is premised on the regulatory proposal, is being discussed in the direction of expanding the scope of PLD to include software.

- Movements in other countries: There are strategic moves to take the global lead in each country, such as the substantial test run of ISO/IEC42001 in Canada and the AI Trust label initiative in Germany.

- New York City: A law has been enacted that requires bias audits, etc. for automated employment decision tools (AEDT), and this trend may spread in the U.S. in the future.

What AI System Audits Should Be

An audit is conducted to visualize an organization's management system, which cannot be viewed objectively by the organization itself, by comparing it with regulations, guidelines, checklists, etc., in order to identify issues. The underlying regulations (rules in the social system) and organizational rules will need to be a combination of two parts for a technology like AI that is changing dramatically in an ongoing manner: a fixed foundation and a variable part, such as an immediate feedback system when an incident occurs. It is also important to balance the appropriate demarcation points of accountability along the supply chain, not to place excessive responsibility solely on the system provider. Individual organizations need to be purpose-oriented and identify areas where rules are necessary, while at the same time paying sufficient attention to those areas that are related to social impact, safety, human life, and human rights. It is also important to ensure a continuous flow of review when it is not possible to immediately determine whether the rules are appropriate or not.

[2] https://www.nist.gov/system/files/documents/2022/12/04/Joint_TTC_Roadmap_Dec2022_Final.pdf

Challenges Specific to AI System Audits

The unique issues of AI system audits include various perspectives that differ from those of conventional system audits, such as how to determine the boundaries of responsibility for increasingly complex and sophisticated operations among mixed stakeholders (for example, how to organize multi-agent AI), whether to target organizations or individual AI systems in the first place, how evidence should be captured, and how optimal knowledge management should be. It is also important to note that the definition of quality itself, which Japan has excelled at in the past, may change.

Given that auditing requires reproducibility (the results do not change no matter who conducts the audit), what is required in auditing AI systems is how to evaluate learning, and it is necessary to be able to distinguish between learning that changes the results and malfunctioning that changes the results. There is still much work to be done to achieve a systematic audit from this perspective.

3. Main comments from the participants

The main discussion topics are as follows.

- **The reality of participatory rulemaking**
  - ✓ It is not yet clear how the private sector is participating in the formulation of regulations in the U.S. and Europe. Even in the U.S., it is recognized that participation by small and medium Enterprises (SMEs) other than GAFAM is still an issue. The Spanish Regulatory Sandbox initiative is also mainly driven by large companies, and the second phase is expected to broaden its base.
  - ✓ EU proposal for AI Act shows a direction that communication with SMEs and NPOs will also be designed in the future.
  - ✓ With the expansion of generative AI, how it is used in end-use and its feedback will become increasingly important in terms of responsibility for use on the part of users and communities. How to design communication with stakeholders will also become important.

- **Steps toward Implementing AI Audits**
  - ✓ Regarding the ISO/IEC42001 test run, it is necessary to have a certain size of company that can run the test, and it would be necessary to target companies of that size without being biased toward industry.
  - ✓ For example, in Germany, based on the business model to be strategically enhanced within the framework of Industry 4.0, the introduction of the system could

be tested from there.

- ✓ There are many examples of companies using GAFAM's services that have become influential tech companies. They could also become test cases for audits.
- ✓ As mentioned above, the very definition of "quality" that Japan has traditionally excelled at may be changing. It is necessary to catch up with the latest trends and discuss them.

- **Direction of rulemaking in Japan**
  - ✓ Japan seems to be good at dealing with gray areas, where people choose not to do something as a matter of common sense, even if there are no clear rules. AI systems change rapidly and will not work if everything is made into rules and documentation is mandatory.
  - ✓ Europe's policy is to go completely hard law. The U.S., while positioning risk management as soft law, is in some respects trying to fill in the gaps in protecting people and society by using soft law to fill in the missing parts based on existing hard law.
  - ✓ Although Japan is moving toward a soft law approach, there is a tendency to follow the guidelines issued by the central government in a hard law-like manner. Even it is soft law, it is important to be careful not to become a disincentive to innovation.

- **Required management and audit structure**
  - ✓ For companies that are going to implement AI systems, the necessary audit structure and requirements should not be a barrier to AI implementation.
  - ✓ It is necessary to ensure that the existing governance and management structure and the audit structure do not duplicate management. It is also necessary to maximize audit productivity while minimizing the burden on the practical side, and to conduct appropriate audits without excesses or deficiencies, from the perspective of both defensive governance and offensive governance that promotes innovation.

- **Human Resources needed for AI Governance**
  - ✓ There are concerns about how to train and secure personnel who can objectively conduct audits, and whether such necessary personnel can be sufficiently recruited.
  - ✓ AI education is often focused on the engineering domain, but traditional methods of studying hardware and software will not catch up. Systematic and bird's-eye view study of System of Systems (SoS) and AI ecosystems, knowledge of management systems to implement them in organizations, and AI auditing skills are also further needed. Human resource development in this area is one of Japan's major challenges.

The 6th Session (Phase III) of the Study Group

Date/Time: January 24th (Tuesday) 15:00-16:00 (On Zoom)

Contents:

- “Recent trends and issues related to AI System Audits” by Hiromu Kitamura (CDLE AI legal group (NEC))
- Questions & Discussion