November, 2020
JDLA

Study Group 'AI governance and its Evaluation'
Report on the Session #3

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 3rd session held on September 25th 2020, under the theme of "Audit and Assurance for AI Governance," Mr. Takashi Akoshima of Japan Digital Design, Inc. talked about "internal audit conducted from a standpoint independent of the business within the company", and Mr. Tomoharu Hase of Deloitte Touche Tohmatsu LLC talked about "external audit and assurance conducted by external auditors from outside the company".

2. Challenges and prospects for internal auditing of AI systems

First of all, Mr. Akoshima gave a topical presentation titled "Internal Auditing of AI Systems - Key Points and Process of Auditing". The opinions expressed in this presentation are from his own perspectives and not the official views of his company.

Focal Points for Auditing AI Systems

While the use of AI is growing, there are no clear standards or guidelines for internal auditing of AI. Therefore, Mr. Akoshima obtained knowledge from system auditing standards and guidelines, papers written by experts, and his research activities at ISACA (Information Systems Audit and Control Association), developed his own frameworks, grasped the overall picture, and proceeded with internal audits by focusing on the parts that fit his company and assessing risks. This presentation particularly focused on evaluating the implementation value and the learning process of AI, which are the characteristics of AI development. The standards and guidelines that he referred to are summarized in Appendix 1.

Points and Process of AI System Audit

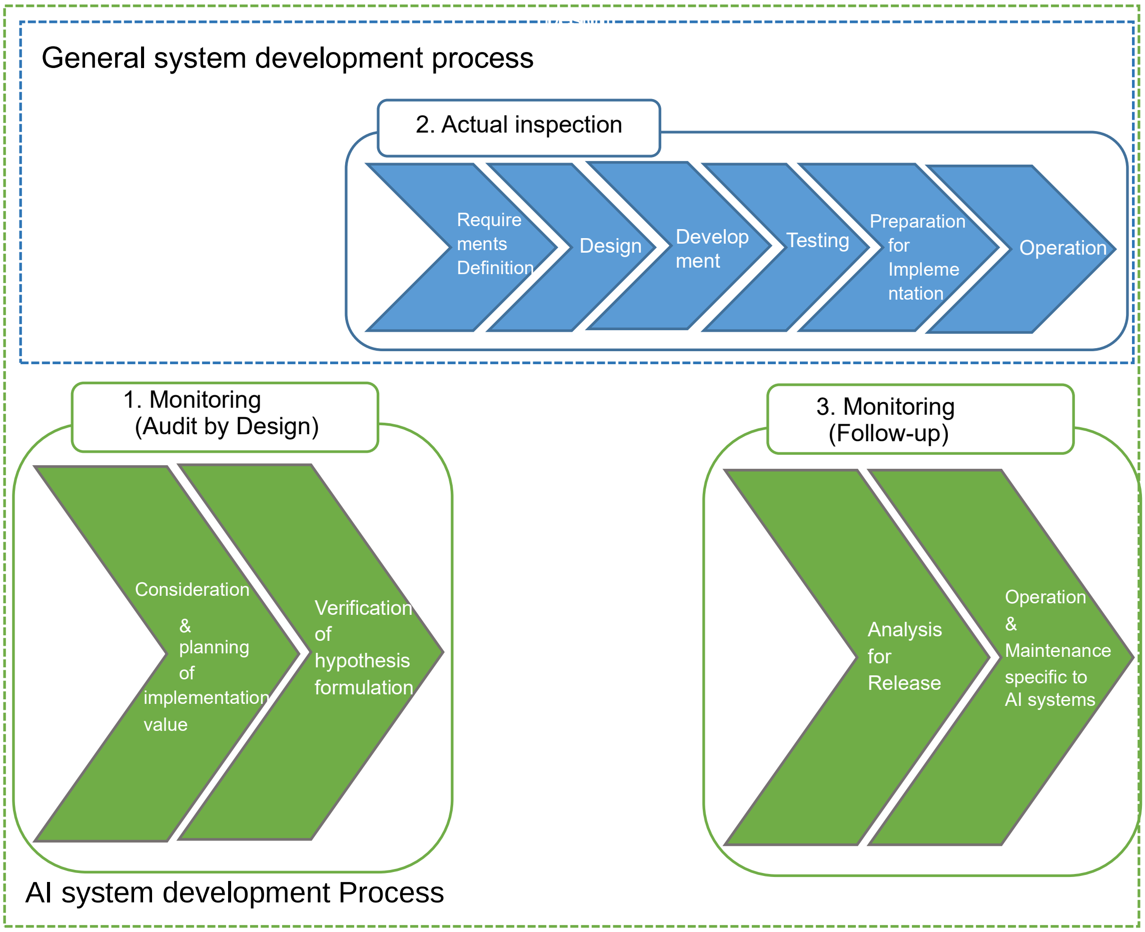
In general, system audits target each of the following processes: "Requirements

Definition, Design, Development, Testing, Preparation for Implementation, and Operation," and evaluation (actual inspection) is conducted by tracing development documents (requirements definition documents, test specifications, etc.) created at specific times in each process.

When it comes to and AI Systems, the "consideration and planning of implementation value and verification of hypothesis formulation" of AI itself is conducted before "Requirements Definition”. In parallel with the "Preparation for Implementation," "Analysis for Release, and Operation and Maintenance specific to AI systems" are conducted. (Figure 1)

As it is difficult to properly evaluate these AI system-specific processes at a specific point in time, continuous monitoring is needed.

Figure1: Flow of AI System Audit


General system development process
2. Actual inspection
Require ments Definition
Design
Develop ment
Testing
Preparation for Impleme ntation
Operation
1. Monitoring (Audit by Design)
Consideration & planning of implementation value
Verification of hypothesis formulation
3. Monitoring (Follow-up)
Analysis for Release
Operation & Maintenance specific to AI systems
AI system development Process

In monitoring phase (Audit by Design), which is conducted before system development and in the planning stage when the value of implementation is being considered, the audit department should audit closely with the business and management divisions, and if there are any issues, they should be addressed as soon as possible. In the hypothesis development and verification phase, it is necessary to clarify the purpose, desired results, and methods of verification, and to confirm that the verification is completed. In the monitoring phase (Follow-up) after system development, it is necessary to analyze the learning status at the time of release, and to confirm that the process of grasping the learning status is incorporated and implemented during operation as well.

In AI system auditing, the departments to be audited differ in each phase. For example, in the monitoring phase (Audit by Design), the business units are the main target, and the business administration division is also included. In the actual audit, the system department becomes the main target, and in the final monitoring phase (Follow-up), the business department becomes the main target.

The key elements of AI system audits can be summarized as shown in Table 1. In each of the processes shown in Figure 1, it is not necessary to go through all the elements, and the audit should be conducted in the context of contents of the AI system and each process.

Table1: Key elements of AI audits

Key elements	Specifications
AI Governance	✓ Basic policies (implementation & development policies, management approval, etc.) ✓ Management system (clarification of authority & responsibility, boundary clarification of ‘white box’ and ‘black box’, accountability, etc.)
AI system development and change management	✓ Planning (formulation of plans, clarification of cost effectiveness, etc.) ✓ Implementation (functional & usage conditions, testing, User Acceptance Test, etc.)
Data Management	✓ Input data (reliability, sufficiency, etc.) ✓ Processing (appropriateness of learning model, verifiability, etc.) ✓ Output data (verification of results, correction of inappropriate results, etc.)
Operation Management	✓ Maintenance & operations (big data, maintenance process considerations, etc.) ✓ Potentiality, Business Continuity Plan (availability, data disposal, etc.)
User Support	✓ Accountability, information provision ✓ User support (promotion of understanding of characteristics, selection of AI use, etc.)
Security Measures	✓ Consideration that the results of the AI system will not harm life, limb, or property and clarification of

	responsibility between AI and human decisions, etc. ✓ Ethics (Respect for individual dignity and autonomy) ✓ Controllability (controllability during development and risk assessment) ✓ Security management (ensuring confidentiality, availability of open source, etc.) ✓ Privacy (privacy protection, infringement) ✓ Intellectual Property (considerations for Intellectual Property Rights)
AI system collaboration	✓ Collaboration between AI systems (Risks in interconnection, etc.)

As mentioned above, the conventional auditing methods are not sufficient for auditing AI systems, and internal auditors of each enterprise need to build an AI audit practice, present examples, and study them together.

3. Challenges and Prospects for External Audit and Assurance of AI Systems

Next, Mr. Tomoharu Hase, Deloitte Touche Tohmatsu LLC, gave a presentation titled "Prospects and Challenges for Assurance Programs for AI Systems". The opinions in this presentation are his own perspectives and not the official views of his firm. Even among audit firms, discussions on AI auditing have just started and there is no unified view on it.

Social demand for AI system assurance

Various social problems are arising in the use of AI. Under such circumstances, there are cases where users feel uneasy about using AI services, and cases where upper management shows concern when companies form alliances with AI systems vendor and try to promote joint business. Therefore, there will be situations where a company will need an assurance from third-party.

Generally, the assurance report is issued after the external auditor has audited the procedures being implemented by the service provider, and the service provider presents the assurance report from the external auditor to the users as necessary. Therefore, it is necessary to design the system by considering what the users demand. Procedures and results that only the auditors and service providers think are good may not necessarily increase the satisfaction of users. Assurance reports on AI systems will lead to reassurance of users, which in turn will promote the use of AI, and will enable AI service providers to develop AI-based services with confidence.

Assurance services provided by audit firms

The typical system-related assurance reports from an audit firm are the SOC reports[1], and there are three types: SOC1, SOC2, and SOC3.[2]

A SOC1 report focuses on the internal control of the service organization over the financial reporting of the user entities. For example, when a user assesses J-SOX [3] for a cloud-based accounting system, it is difficult for the service organization to respond to all requests for an audit for each user company. Therefore, the service organization will reduce the burden of J-SOX compliance by obtaining SOC1. The scope of SOC2 includes areas outside of financial reporting, such as security, availability, processing integrity, confidentiality, and privacy.

The object of assurance in SOC is not to guarantee that users will not suffer any disadvantage due to the use of the system, but to express an assurance opinion on the effectiveness of internal control. In other words, it assures that the contents of the statement are correct, that the service organization's internal control is appropriately designed, being operated effectively for the specified period (six months to one year). As for the assurance of AI, Mr. Hase mentioned that it is highly likely that audit firms will provide assurance on the effectiveness of internal controls, but it is difficult to assure that AI will not discriminate or cause any accidents.

What does it take to assure AI systems?

In order to assure AI systems, it is important to have an appropriate setting and agreement on the expected assurance level of "what and how much is assured" among the three parties: the party to be assured (AI service companies, etc.), the party to assure (audit firms, etc.), and the party to use the assured results. One of the reasons for this is that when AI system service providers receive assurance, they not only incur audit costs, but also increase their workload by creating and storing documents necessary for the audit. Secondly, there is a risk of not achieving the expected cost-effectiveness at the time of introduction of AI services due to excessive monitoring in normal operations to meet the excessive requirement level. Therefore, it is essential to set the expected assurance level appropriately.

In establishing an assurance system for AI systems, it is necessary to create evaluation criteria and assurance standards, as well as professionalism, independence, and quality

[1] The American Institute of Certified Public Accountants (AICPA) has defined a framework for internal control assurance reporting for service organizations and cybersecurity as System and Organization Controls (SOC).

[2] SOC1 and SOC2 are the main ones, and SOC3 is rarely issued; SOC1 and SOC2 have Japanese, American, and international standards, and the names of the standards are different even for the same SOC.

[3] An internal control reporting system. It is a system that requires listed companies to submit an internal control report along with their annual securities report.

control systems to prevent variations in procedures and results among audit firms. In addition, in order to ensure comparability of reports, it is necessary to standardize the content and format of reports.

Currently, assurance standards and report templates are provided by SOC though, there are no system evaluation criteria that are specific to AI systems yet. One of the difficulties in establishing AI system evaluation is that the timing of evaluation differs from that of normal system development. For example, if the collection of training data or bias is pointed out after the completion of an AI system, the provider company is unable to respond. In order to properly evaluate the system, it is difficult to determine from which stage the auditing firm should be involved, and it is highly likely that a new scheme such as cooperation between internal and external audits will be necessary.

As another possible option, there are ways to use existing frameworks for assurance process. One of the existing frameworks to be considered is SOC2+, which is an assurance report that can include not only the evaluation criteria defined by SOC2, but also the criteria of other external organizations. The advantage is that it only requires the creation and addition of system evaluation criteria for AI. The disadvantage is that it is necessary to comply with the SOC2 evaluation criteria as well, which raises the hurdle of obtaining an assurance report.

4. Discussion points in the question and answer session

The third session focused on auditing AI Systems, and discussion were held from the perspective of internal and external auditing. Based on the topics presented, the following questions and answers were raised.

Internal Audit

- ✓ The criteria to be evaluated by internal audit is whether the process was properly planned and executed. In the evaluation at the planning stage of considering the value of AI implementation, it is necessary to evaluate the appropriateness of the process, such as whether it has been discussed at the board of directors meeting. Although it is not possible to evaluate the content of the process at the stage of hypothesis formulation and verification, it is preferable to view the evaluation plan and results, which include numerical and other evaluation criteria, and evaluate them based on the criteria of whether the process has been incorporated, implemented, and evaluated.
- ✓ If the risk of the AI service is low, the management department is considered to be accepting of the risk and no audit is conducted. Basically, internal auditing is an evaluation process that is close to the business and management departments.

- ✓ It is more difficult to draw a line to determine that there is no problem with AI systems than with non-AI systems. In terms of auditing AI systems, the focus should be on the process, but the sufficiency of the process depends on the risk, the company, and the AI system, and cannot be generalized. Until now, system auditors have only had to evaluate systems themselves, but given that AI is embedded in business systems, it will be necessary to evaluate AI in business audits. Therefore, it is necessary for auditors who conduct business audits to acquire the ability to audit AI and to strengthen their professional capabilities.
- ✓ It is difficult to evaluate an AI system by itself, and it is necessary to evaluate it as a business process. It is important to think of the entire process as a way to reduce risk.
- ✓ In general, external audits do not look at cost effectiveness. On the other hand, internal audits need to confirm whether the business generates value and how management judges its future potential.
- ✓ It would be good for companies using AI systems to start by doing a risk assessment to see how big the risk of their AI is.

External Audit & Assurance

- ✓ The evaluation of AI systems depends not only on the AI service providers, but also on the AI service users. The SOC report describes the controls that need to be verified on the user side.
- ✓ When considering the assurance of AI, focusing on the process will result in high costs; it is necessary to take AUP [4] into consideration. However, in some cases, the satisfaction of the end user does not increase because there is no assurance opinion in AUP. In light of this, it is necessary to consider whether to take AUP or assurance.
- ✓ With regard to the 2C (to consumer) in B2B2C (Business-to-business-to-consumer), rather than a guarantee, the basis will be to explain what should be secured by the users and have them agree to it, and users will be expected to use the service after agreeing to the terms of use, etc.
- ✓ It is difficult to develop a “one size fits all” audit framework which satisfies all cases. It is necessary to proceed on a case-by-case basis, and this point needs to be worked out in the study group.

[4] Agreed-Upon Procedures. One in which a practitioner is engaged to issue a practitioner's report of findings based on agreed-upon procedures applied to subject matter.

Fairness, Bias in AI

- ✓ Regarding the assurance of fairness in AI, the level and definition of fairness and ethics differ from country to country. There is a great deal of room for discussion as to whether the evaluation criteria for assurance should be aligned with ethics in Japan or globally.
- ✓ Biases caused by AI systems varies by case study and field. Consideration should be given early on whether those biases are ethically permissible. Audits need to look at whether there is a process in place to check during the training phase, whether follow-up after release is being evaluated, and whether the system is overfitted.

Roles of AI vendors

- ✓ From the standpoint of a vendor, it is unclear whether it is possible to design a system that assures fairness and other ethical perspectives for specific usage scenarios. It is necessary to consider how to address the ethical challenges posed by AI as vendor side.
- ✓ In the case of supervised learning, it is difficult for the vendor to evaluate the data quality, as the business department is required to determine whether the given data and labeled data are appropriate or not. It is necessary to consider who should play the role of assessing data quality.

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

Written by Yuki Kiyomi
Translated by Michiko Shimizu

<Outline of the 3rd Session of the Study Group>
Date & Time: Tuesday, September 25, 2020, 16:00-18:00 (Zoom)
Agenda:

- Topical presentations:
  "Internal Auditing of AI Systems - Key Points and Process of Auditing” provided by Mr. Takashi Akoshima (Japan Digital Design, Inc.)
  "Prospects and Challenges for Assurance Programs for AI Systems" provided by Mr. Tomoharu Hase (Deloitte Touche Tohmatsu LLC)
- Question and answer session / discussion

Appendix 1: AI system-related standards and guidelines used as reference

(organization, name of reference material, date of publication, URL)

1. AI Network Society Promotion Council, Draft Guidelines for AI Development for International Discussion, July 28, 2017, https://www.soumu.go.jp/main_content/000499625.pdf
2. Institute of Information and Communications Policy, Ministry of Internal Affairs and Communications, Draft Principles for AI Utilization, July 31, 2018, https://www8.cao.go.jp/cstp/tyousakai/humanai/4kai/siryo1.pdf
3. Ministry of Economy, Trade and Industry, System Auditing Standards, April 20, 2018, https://www.meti.go.jp/policy/netsecurity/downloadfiles/system_kansa_h30.pdf
4. Ministry of Economy, Trade and Industry, System Management Standard, April 20, 2018, https://www.meti.go.jp/policy/netsecurity/downloadfiles/system_kanri_h30.pdf
5. Ministry of Economy, Trade and Industry, Study Group on Contract Guidelines for the Use of AI and Data, December 2019, https://www.meti.go.jp/press/2019/12/20191209001/20191209001-1.pdf
6. Integrated Innovation Strategy Council, Social Principles of Human-Centric AI, March 29, 2019, https://www8.cao.go.jp/cstp/aigensoku.pdf
7. Center for Financial Information Systems, System Auditing Standards for Financial Institutions, March 2019
8. Doubunkan Publishing, Yuji Shimada, Understanding System Auditing in Practice, 3rd Edition, January 25, 2019
9. The Institute of Internal Auditors, GLOBAL PERSPECTIVES AND INSIGHTS Artificial Intelligence – Considerations for the Profession of Internal Auditing
10. The Institute of Internal Auditors, GLOBAL PERSPECTIVES AND INSIGHTS The IIA's Artificial Intelligence Auditing Framework Practical Applications, Part A
11. The Institute of Internal Auditors, GLOBAL PERSPECTIVES AND INSIGHTS The IIA's Artificial Intelligence Auditing Framework Practical Applications, Part B
12. Future of Life Institute, ASILOMA AI PRINCIPLES, https://futureoflife.org/ai-principles/?cn-reloaded=1
13. OECD, Council Recommendation on Artificial Intelligence, adopted on 22 May 2019, https://www.soumu.go.jp/main_content/000642217.pdf
14. National Institute of Advanced Industrial Science and Technology, Machine Learning Management Guidelines, June 30, 2020, https://www.cpsec.aist.go.jp/achievements/aiqm/