Case Study of Human Resources Development for AI Risk Management Using RCModel

ITO Hirohiko, ITO Chihiro

Abstract

As AI has become more widely used in recent years, AI governance to ensure proper use of AI has been drawing attention. Some regulations related to AI governance require appropriate risk management by people. Therefore, developing human resources who can be responsible for the risk management of AI services is expected to become an important issue. Since 2021, NEC has been conducting joint research with the University of Tokyo on how to develop AI-specialized human resources for AI risk management using the RCModel, a tool developed by the University of Tokyo. This paper will provide an overview of the human resource development program implemented in the joint research as well as the results and the achievements thereof and present the future prospects of the program.

Keywords

AI governance, AI ethics, human resource development, RCModel, risk management

1. Introduction

AI governance has attracted attention in recent years. Over the past 10 years, the use of AI has spread explosively in society and as a result, inappropriate use of AI and discriminative determination by AI from unnoticed bias have become problems. To counter such risks posed by the use of AI, policies and guidelines for the development and use of AI have been developed at various levels, from companies to governments. Currently, in addition to the development of those policies and guidelines, attention is now focused on how governance should be to ensure their implementation.

In particular, the development of human resources responsible for risk management is expected to become an important issue in AI governance. EU AI regulations—a hot topic lately—require AI that is considered to be high risk to be monitored by people and also require the people who monitor AI to sufficiently understand the limitations and possible risks of AI. However, AI-specialized human resource for risk management has not been sufficiently considered, even in the Digital Skill Standard[1] published by the Information-technology Promotion Agency (IPA) and the Ministry of Economy, Trade and Industry in Japan.

Considering this situation, NEC has been working together with the University of Tokyo on joint research to develop AI-specialized human resources responsible for risk management since 2021. Specifically, it has considered human resource development programs in collaboration with the NEC Academy for AI, a one-stop service for DX human resource development, using the Risk Chain Model (RCModel) developed by the University of Tokyo.

Next, this paper provides an overview of the RCModel followed by information about the NEC Academy for AI. Then, an overview of the human resource development program conducted as a trial and the results thereof are described, ending with a description of the future outlook for the program.

2. RCModel: an AI Risk Management Support Tool

The RCModel, which is being developed by the University of Tokyo[2], is a tool to identify the risks of AI services that are planned, developed, and operated by service
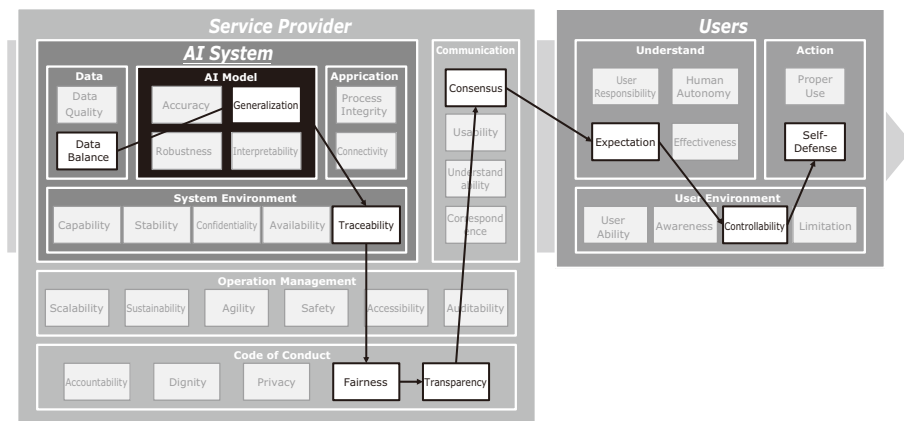
Fig. 1 Example of use of RCModel.

providers themselves, to consider how to minimize those risks, and to explain this to third parties.

The RCModel organizes risk scenarios and responses (controls) for specific AI service cases from an overview of the AI (such as the values of service and objectives to be achieved and system configurations), thereby summarizing where the achievement of the values and objectives is hindered. The values and objectives here refer to the reasons for introducing AI, such as improving productivity, preventing accidents, and improving customer satisfaction. The risk scenarios refer to cases where the values and the purpose for which AI was introduced might be hindered, such as when the accuracy of AI is degraded or when the AI's behavior differs depending on the person. It should be noted that there are a wide variety of possible risk scenarios to consider, and there is no single correct way to respond to them.

The RCModel has a hierarchical structure (AI model, AI system, service provider, and user) based on 38 structural elements (predicted performance, data quality, fairness, user's responsibility, etc.) that summarize particulars that should generally be considered for AI. When considering responses to different risk scenarios, the RCModel is used to review the specific tasks involved and their sequence, highlight the components relevant to each task, and visually link these highlighted components, creating a chain-like structure.

For example, in the example shown in **Fig. 1**, the relevant items are chained in the following order: Data Balance (data distribution) → Generalization (judgment that is less biased towards specific cases) → Traceability (characteristics that allow for post factum verification of the AI service) → Fairness (fairness throughout the service) → Transparency (disclosure of necessary information related to the AI service) → Consensus

(alignment of understanding with users) → Expectation (understanding of the expected accuracy of the AI service) → Controllability (control on the part of the user) → Self Defense (user's own protection). Although we will not detail those items in this paper, the connections of the risk chain show that responses to a certain risk scenario are to be carried out in cooperation with the organizations and personnel involved in the layers of the hierarchy that include the individual items. This example shows that responses to the risks span multiple organizations and stakeholders.

In this way, the RCModel has the distinction of being equipped even with a framework specifically for describing risk responses. It also works well with other frameworks used to consider the risks of AI services, such as the Digital Ethics Compass[3], so NEC is taking note of the RCModel.

3. NEC Academy for AI

Against the backdrop of the development of digital human resources being required across Japan, NEC provides human resource development programs at its NEC Academy for DX. In particular, NEC launched the NEC Academy for AI in 2019 as a program aiming to develop human resources who specialize in AI and data-related areas.

The NEC Academy for AI provides a variety of programs with different content and durations, but its longest-term program is the one-year program that serves as an entrance course where students acquire knowledge and practical experience through training, simulated exercises, and practice (on-the-job training (OJT) on actual projects) with mentors. The NEC Group and user companies have students they select (academy students)
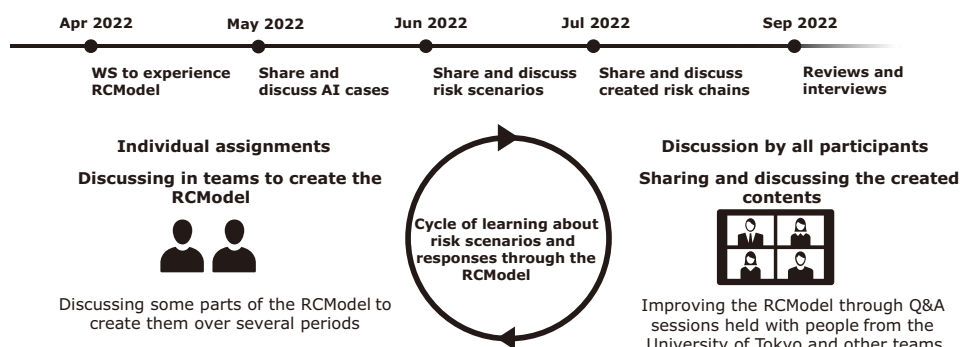
Fig. 2 Overview of the human resource development program.

to participate in the entrance course to develop them as leaders who will lead their digital transformation (DX) departments in the future.

4. Program to Develop Human Resources Responsible for AI Risk Management

4.1 Overview of the human resource development program

The human resource development program at this time was held in a manner to provide four rounds of workshops (WS) to four students in the NEC Academy for AI's entrance course in approximately three months from April to July 2022. The first round was a guidance session that provided an overview (**Fig. 2**).

The academy students were divided into two teams of two in accordance with the industry of their employer, and each team was asked to consider a hypothetical AI service as a case study in an individual assignment. The teams were asked to consider the risks of the hypothetical AI service established by each team and responses to the risks, and then they made presentations of the considered risks and responses thereto to related parties from the University of Tokyo and other team and held discussions as part of the workshop, thereby improving the assignment results.

The reason the human resource development program was provided in the workshop format was because there were two main expectations.

The first is the expectation that the RCModel help us notice risks from different perspectives (positions, fields of view, and points of view) and consider more appropriate responses. In recent years, it is not uncommon for AI services to be implemented and operated by multiple companies. In that case, a variety of expertise (such as detailed knowledge business and industry practices) is required for risk management of AI services. In addition, such specialized knowledge is undocumented and is tacitly known in many cases. We thought that by expressing the know-how related to the implementation and operation of AI services as explicit knowledge in the framework of the RCModel and by having both teams learn the know-how from each other through discussions in the workshops, risk management skills would improve.

The second is the expectation that the RCModel will help us recognize a variety of stakeholders and realize the need to build consensus among them. The RCModel is used to categorize AI systems, system providers, users, and stakeholders. The connections of the risk chain emphasizes the need for these stakeholders to collaborate when responding to risks. This also shows that related stakeholders must agree on how to respond. We thought through workshop discussions we could consider what kind of stakeholders there are and how they can share roles.

4.2 Evaluation of the human resource development program

After finishing the workshops, we interviewed students and mentors regarding their evaluations of the human resource development program. We obtained the following comments from the students.

Recognizing and considering risks and responses that would otherwise not be recognized or considered helps improve risk management skills.

Table 1 summarizes comments about considering risks and responses. These comments indicate that the RCModel is a tool that helps you notice risks that you would otherwise not notice yourself and helps you consider response flows that are difficult to imagine yourself. The comments also showed that discussion with more than one person in the workshop format was expected to increase the effectiveness of the tool. As stated in section 4.1, the RCModel records both the tacit knowledge

Table 1 Comments on considering risks and responses.

Comment	Commented by
We could share a common understanding of the risk scenarios and responses even with people who had no expertise in AI.	Student A
I am regularly engaged in system development and data science work. By using the RCModel, it becomes possible for me to consider risks and responses from the perspectives of someone on-site.	Student B
By considering everything in teams, we were able to supplement each other's knowledge from different perspectives.	Student C
I felt the use of the RCModel enabled us to also use it as a flowchart of business operations including the BCP perspective.	Student D
Consideration with the RCModel methodology could enable us to notice risks and responses we may not otherwise notice through our own experience.	Mentor

Table 2 Comments on agreements with stakeholders.

Comment	Commented by
The use of the RCModel would help us make presentations to the public about AI risk control.	Student A
Through the workshops, I felt that it is important to clarify the roles of each stakeholder in using AI.	Student C

of AI service risks and the know-how on how to respond to the risks as explicit knowledge. As a result, we can see that this tool supports learning about other people's perspectives. These comments indicate that by discussing specific cases with a variety of stakeholders, we can expect to examine risks from various perspectives and consider more appropriate responses.

Recognizing the importance of role sharing between stakeholders and encouraging consensus building with the RCModel as a common framework

Table 2 summarizes comments on agreements with stakeholders. These comments suggest that consideration of AI service risks and responses thereto through the use of RCModel allows the importance of role sharing between stakeholders to be recognized. In addition, the use of RCModel as a common framework is expected to encourage negotiations and consensus building with diverse stakeholders which are essential for risk management of AI services.

5. Conclusion

This paper presented a review of the human resource development program collaborated on with the NEC Academy for AI, NEC's DX human resource development institution, using the RCModel. As stated at the beginning of this paper, the need for human resources who can manage the risks of AI services is also increasing alongside society's growing attention to AI governance.

The trial at this time revealed that the program in the workshop format using the RCModel allows for the learning of risks and responses that may not otherwise be recognized or considered by an individual. It also encourages consensus-building by highlighting the importance of role sharing between stakeholders. We believe that these can serve as useful methodologies for developing human resources responsible for managing AI risks.

NEC will proactively utilize this achievement in the programs of the NEC Academy for AI.

References

1) Ministry of Economy, Trade and Industry: Digital Skills Standards, August 2023 (Japanese)
https://www.meti.go.jp/policy/it_policy/jinzai/skill_standard/main.html
2) Technology Governance Policy Research Unit, Institute for Future Initiatives, The University of Tokyo: Risk Chain Model (RCModel) Guide, Ver. 1.0, June 2021 (Japanese)
https://ifi.u-tokyo.ac.jp/wp/wp-content/uploads/2021/07/RCM_210705.pdf
3) DDC:The Digital Ethics Compass
https://ddc .dk/tools/toolkit-the-digital-ethics-compass/

Authors' Profiles

ITO Hirohiko
Professional (R&D Open Innovation Coordinator)
Global Innovation Strategy Department

ITO Chihiro
Manager and Lead Planner for DX Learning Strategies
AI Analytics Department

The details about this paper can be seen at the following.

Related URL:

To Realize an Open and Trustworthy AI Society-Discussion on AI Ethics by The University of Tokyo and NEC (Japanese)
https://wisdom.nec.com/ja/feature/ai/2023042101/index.html
NEC Academy for AI Human Resources Development Program; "Webinar on the Application of the Risk Chain Model: Risk Chain Model to Practice" (Japanese)
https://ifi.u-tokyo.ac.jp/event/14233

Information about the NEC Technical Journal

Thank you for reading the paper.
If you are interested in the NEC Technical Journal, you can also read other papers on our website.

Link to NEC Technical Journal website

Japanese	English

Vol.17 No.2 Special Issue on Revolutionizing Business Practices with Generative AI
— Advancing the Societal Adoption of AI with the Support of Generative AI Technologies

Remarks for Special Issue on Revolutionizing Business Practices with Generative AI
Approaches to Generative AI Technology: From Foundational Technologies to Application Development and Guideline Creation

Papers for Special Issue

Market Application of Rapidly Spreading Generative AI
NEC Innovation Day 2023: NEC's Generative AI Initiatives
Streamlining Doctors' Work by Assisting with Medical Recording and Documentation
Using Video Recognition AI x LLM to Automate the Creation of Reports
Understanding of Behaviors in Real World through Video Analysis and Generative AI
Automated Generation of Cyber Threat Intelligence
NEC Generative AI Service (NGS) Promoting Internal Use of Generative AI
Utilization of Generative AI for Software and System Development
LLMs and MI Bring Innovation to Material Development Platforms
Disaster Damage Assessment Using LLMs and Image Analysis

Fundamental Technologies that Enhance the Potential of Generative AI
NEC's LLM with Superior Japanese Language Proficiency
NEC's AI Supercomputer: One of the Largest in Japan to Support Generative AI
Towards Safer Large Language Models (LLMs)
Federated Learning Technology that Enables Collaboration While Keeping Data Confidential and its Applicability to LLMs
Large Language Models (LLMs) Enable Few-Shot Clustering
Knowledge-enhanced Prompt Learning for Open-domain Commonsense Reasoning
Foundational Vision-LLM for AI Linkage and Orchestration
Optimizing LLM API usage costs with novel query-aware reduction of relevant enterprise data

For AI Technology to Penetrate Society
Movements in AI Standardization and Rule Making and NEC Initiatives
NEC's Initiatives on AI Governance toward Respecting Human Rights
Case Study of Human Resources Development for AI Risk Management Using RCModel

NEC Information

2023 C&C Prize Ceremony

NEC Technical Journal

Vol.17 No.2
June 2024

Special Issue TOP