AI Ethics Impact Assessment based on requirement engineering

Izumi Nitta
Fujitsu Limited
Kawasaki, Japan
nitta.izumi@fujitsu.com

Kyoko Ohashi
Fujitsu Limited
Kawasaki, Japan
ohashi.kyoko@fujitsu.com

Satoko Shiga
Fujitsu Limited
Kawasaki, Japan
shiga.satoko@fujitsu.com

Sachiko Onodera
Fujitsu Limited
Kawasaki, Japan
sachiko@fujitsu.com

***Abstract* —This paper proposes a methodology for evaluating the ethical impact of AI systems on people and society based on AI ethics guidelines. In recent years, the ethical impact of AI has been recognized as a social issue, and countries and organizations have formulated principles and guidelines on AI ethics, and laws and regulations will be enforced in Europe. Because these principles and guidelines are written in terms of philosophy and law, AI service providers, developers and business users have a challenge that how they should practice the principles and guidelines to their AI systems. To address this challenge, we firstly analyzed cases of ethical problems caused by AI in the past and hypothesized that ethical problems could be linked to interactions between components of AI systems and stakeholders related to AI systems. Then, based on this hypothesis, we developed a methodology to comprehensively extract the ethical risks that the AI system poses. This methodology consists of two approaches. The first approach is to develop an AI ethics model that embodies ethics guidelines as necessary requirements for ethical AI systems and correlates these requirements with interactions. The second approach is an impact assessment process that uses the AI ethics models to extract ethical risks for individual AI systems. In this paper, we show the details of the methodology and show the results of initial experiments to verify the validity of the hypothesis and the ease of the impact assessment process.**

Keywords — AI ethics, AI governance, responsible AI, impact assessment, risk-based approach

I. Introduction

The ethical impact of AI has become recognized as a social issue, and ethical principles and guidelines that provide basic principles for the spread of responsible AI are being developed [1] [2] [3] [4] [5] [6]. In Europe, the European Commission published a draft AI regulation (Artificial Intelligence Act) [7]. The Artificial Intelligence Act proposal categorizes the manipulation of people's subconscious minds, the use of social scoring, remote biometrics for law enforcement purposes in public spaces, and the like as "AI that create an unacceptable risk" and prohibits this type of AI. It also lists the use of AI in personal biometrics and classification, and its application to critical infrastructure as "high-risk AI," and imposes a number of obligations for its use in these fields, with significant fines for violations. In the United States, a bill known as the "Facial Recognition and Biometric Technology Moratorium Act" [8], which prohibits federal officials from using facial recognition technology, has been proposed. In addition, the city of San Francisco has also banned the use of facial recognition technology by police.

In recent years, the movement from ethical principles to practice has become active. Aiming to comply with ethics guidelines, the development of technology to address ethical issues related to the fairness of AI is moving forward. Several standards have been defined regarding fairness. For example, for the sensitive attributes such as race or gender for which fairness is being considered, there is one approach that equalizes the probability of the potential outputs provided by the AI (for example, in the case of recruitment AI, the hiring and rejection of candidates), and another that equalizes the probability of AI output being adopted in correct data. Machine learning algorithms that follow these various standards are being studied extensively [9]. In addition, a technique [10] that considers fairness in relation to multiple sensitive attributes such as gender and race, and a concept for an interface design for end users to judge and deal with fairness in AI have also been proposed [11].

Meanwhile, efforts have begun to assess the ethical implications of AI systems. The Canadian government has issued guidance on Automated Decision-Making [12] and provided a tool to assess the impact of algorithms on decision-making systems (the Algorithmic Impact Assessment Tool). The Algorithm Impact Assessment toolkit [13] from the Ada Lovelace Institute provides an algorithmic impact assessment for healthcare AI. Floridi et al. provided guidance on the conformity assessment of AI systems based on the European Artificial Intelligence Act proposal [14]. NIST also categorizes AI biases and reports challenges and guidance [15].

In this paper, we consider ethical issues and their causes caused by AI as ethical risks. A framework for assessing ethical risks has also been proposed. The risk chain model [16] provides a framework for AI service providers to consider risk assessment and control of their AI services. The relationship between risk scenarios and risk factors in the AI system is visualized, and the examination of risk control is enabled on the basis of the result.

The purpose of our research is to provide a methodology for AI service providers, AI developers, and business users who are non-AI ethics experts to, with respect to their AI systems, evaluate the ethical impact of AI systems based on AI ethics guidelines, and to recognize where and what risks may occur in AI systems.

Since AI principles and guidelines are written in the language of law and philosophy, for non-AI ethics experts, reading and comprehending these principles and guidelines, and putting them into practice in their own AI systems can be burdensome. Even when using the various impact assessment frameworks discussed above, it is likely that knowledge of AI ethics and past case analysis will be required to apply them to their own AI systems.

In order to address the issue, we hypothesized that systematizing ethical problems caused by AI in the past would enable conducting impact assessment procedurally for various AI use cases. From the survey results, we derived the following hypothesis that ethical risks are mapped to the interactions between the components of an AI system and the stakeholders directly or indirectly involved with the AI system.

Based on this hypothesis, we propose a methodology “AI Ethics Impact Assessment” for comprehensively identifying ethical problems caused by AI by associating the requirements for responsible AI described in the ethics guidelines with the interactions that appear in AI systems. The methodology consists of two approaches based on the requirements engineering approach.

Approach 1: Building an AI ethics model. Embody written ethical guidelines as requirements necessary for AI systems to be ethical, and map these requirements to interactions. Ethical risk can be treated as a situation contrary to the requirements associated with the interaction.

Approach 2: Impact assessment process. The process of extracting ethical risks for individual AI systems using an AI ethics model is presented.

Using the proposed methodology, AI developers, providers, and users without expertise in ethical guidelines can assess ethical risks at each stage of the AI lifecycle, from planning to development, operation, and retirement.

In this paper, we show the details of the AI ethics model and the AI ethics impact assessment, conduct verification experiments on the validity of the hypothesis on interaction and the ease of the impact assessment process, and discuss the effectiveness of and issues concerning the proposed methodology.

Section II analyzes and hypothesizes on cases of past ethical problems, Section III describes procedures for AI ethical impact assessment using the AI ethics model with the proposed methodology, Section IV describes verification experiments on and considerations regarding the proposed methodology, and Section V describes conclusions and future issues.

II. HYPOTHESIS ABOUT ETHICAL ISSUES OF AI

We hypothesized that if we could systematize how ethical issues arise from AI in several patterns, it would be relatively easy for stakeholders who understand the specifications and use cases of AI systems to conduct impact assessments. Accordingly, we analyzed past ethical issues and developed a hypothesis that led to systematization.

A. *How can ethical risks be handled?*

Examples of past ethical problems were observed and whether there were common patterns in the way the ethical problems occurred was examined by the following procedure.

- First, the use cases in which past ethical issues occurred are represented by the components and stakeholders of the AI system and the interactions between them. The visualization of this is called a system diagram.
- Second, it maps where the ethical issues occur in a system diagram.

As a past example, loan screening AI is taken up. The loan screening AI determines whether a loan is approved or not based on the attributes, transaction history, and credit score of the loan applicant. The loan officer makes a final decision based on the AI results and responds to the applicant. The AI model features past loan applicants' attributes, transaction histories, and credit scores, and learns repayment performance as a teacher label. In the past, it has been pointed out that the results of loan screening AI are biased by gender.

Fig. 1 shows where the ethical risks of loan screening AI lie. The diagram shows a simplified AI system, with loan officers and loan applicants as stakeholders. Arrows connecting the AI system's components and stakeholders represent interactions.

The ethical risks that can be mapped to each of the following four interactions are shown:

1. From training data to AI models: Gender bias is learned from training data and reflected in AI model.

2. From AI model to output: There is gender bias in the results obtained using AI model.

3. From loan officer to loan applicant: The loan officer makes final decisions heavily depending on the AI results.

4. From loan applicant to loan officer: The loan applicant makes an objection to the loan officer.

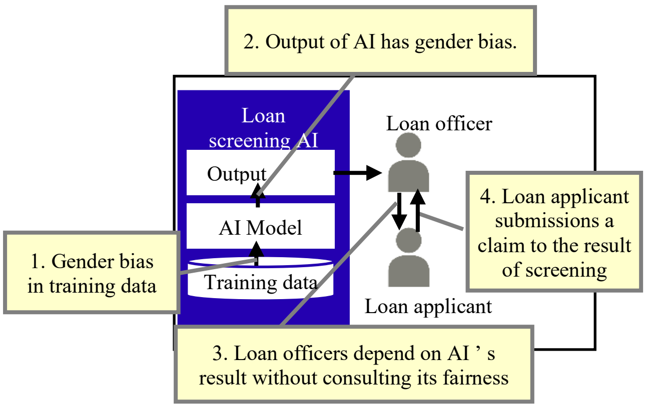


Fig. 1: An example of ethical risks on loan screening AI

B. *Hypothesis*

We conducted an analysis of past ethical issues by AI and developed the following hypothesis.

Hypothesis: Ethical risks relate to the interactions between the components of an AI system and the stakeholders that are directly or indirectly involved with the AI system.

In addition, the analysis of the cases shows that the components of the AI system and stakeholders can be patterned according to their roles. For example, components of an AI system may include training data, machine learning processing, AI models, inference results, etc. Stakeholders may include developers, service providers, data providers, business users, end users, judgment targets, observers, etc. It is considered that by patterning in this way, the above hypothesis can be applied to various AI use cases.

III. PROPOSED METHODOLOGY

This section describes the AI ethical impact assessment that we propose. Based on the hypothesis on ethical risks of AI derived in the previous section, this methodology analyzes where and how the ethical risks that may occur by the use of AI systems appear. The analysis results are described on a system diagram illustrating the AI system, stakeholders, and interactions between them. This methodology can be carried out procedurally in the case of having knowledge about the AI system to be evaluated, and enables reliable evaluations in a realistic time period.

A. AI ethics model

First, we describe the AI ethics model required to extract AI ethical risks. An AI ethics model comprehensively indicates the characteristics that an ethical AI system should have. Individual characteristics included in an AI ethics model are called AI ethical characteristics. An AI ethics model organizes and defines its characteristics hierarchically.

In this context, the AI system is an IT system that uses technologies related to AI. It is assumed to include not only AI models generated using machine learning and statistical analysis, but to be a system in which processing components required for IT systems are combined as well. In addition, stakeholders directly or indirectly involved in AI systems are included in the AI ethics model.

The range of AI ethical characteristics of the AI ethics model shall be the range indicated in the ethical guidelines. We believe that ethical guidelines can be embodied in requirements that AI systems must have. This is based on the concept of requirements engineering. The User Experience (UX) Quality Model [17] combines a top-down approach with a bottom-up approach to comprehensively collect and organize UX quality characteristics. Specifically, the UX quality model is expressed in four levels: the upper two levels are expressed using the definition of SQuaRE [18], and the lower two levels are embodied using the results of UX evaluation by the user. Through this approach, a practical model can be constructed by comprehensively expressing quality characteristics in the upper levels, and preventing excessive embodiment in the lower levels. We used this concept to construct an AI ethics model.

The AI ethics model is also expressed in a four-level structure. In the top-down approach, the text of the ethical guidelines is structured in the upper two levels. Next, in the bottom-up approach, regarding ethical problems obtained from the analysis of AI system use cases [19] and AI system cases that caused problems in the past from AI incident database [20] published by Partnership on AI, a non-profit organization that promotes AI, the characteristics that can be achieved by solving the problems are extracted as AI ethical characteristics, and these are embodied in the lower two levels, and guidelines and AI ethical characteristics are made to correspond to each other.

Furthermore, for each AI ethical characteristic, the kind of interaction (for example, AI ethical characteristics related to AI fairness include handling interactions from AI outputs to business users, etc.) to satisfy the characteristic is extracted and correlated. In this case, a plurality of interactions may correspond to one AI ethical characteristic, and one interaction may correspond to a plurality of AI ethical characteristics. This prevents interactions and AI ethical characteristics from becoming incompatible. These approaches provide the following advantages:

- By basing the approaches on the ethical guidelines, it is possible to ensure completeness in the sense of compliance with the ethical guidelines.
- Because interactions in AI systems correspond to AI ethical characteristics, and guidelines correspond to the upper levels of AI ethical characteristics, it is possible to clarify which part of the guidelines an interaction should correspond to. This makes it easier to take measures such as technical solutions or operational measures.

Fig. 2 shows an excerpt of an AI ethics model. This example applies Trustworthy AI [2] from the EU's AI HLEG to the underlying ethical guidelines. Trustworthy AI consists of seven requirements and assessment lists for each requirement. In this model, seven requirements are specified in the uppermost level. The lowermost level is associated with interactions.

For example, an AI ethical characteristic called "Maintenance of social trust" is associated with an interaction between business user and judgment target. A situation in which the target system violates this requirement constitutes an ethical risk. In the case of the loan screening AI mentioned in the previous section, this requirement allows "the loan officer makes final decisions heavily depending on the AI results" to be extracted as an ethical risk.

With this approach, it is possible to conduct impact assessments for various use cases by creating a model once per guideline.

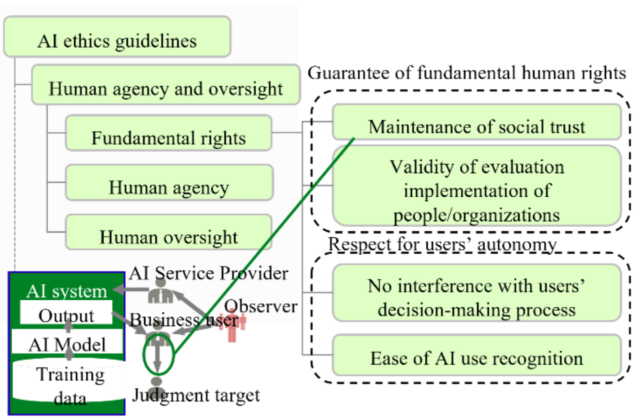


Fig. 2: Excerpt of the AI ethics model

B. *Risk extraction procedure*

Using the AI Ethics Model, we describe how AI Ethical Impact Assessment is performed along the overall diagram shown in Fig. 3. The procedure consists of three steps.

1) Step 1: First, with regard to the AI to be evaluated, create an AI system diagram based on AI specifications and use case information. A system diagram (Fig. 3, Step 1) is a diagram in which stakeholders related to the components (data, AI model, etc.) of an AI system are arranged and their interactions are shown by arrows.

2) Step 2: Next, AI ethical characteristics corresponding to all interactions in the system diagram are extracted using the AI ethics model (Fig.3, Step 2). Step 2 can be mechanically performed using the interaction extracted in Step 1 as an input.

3) Step 3: Finally, regarding each extracted AI ethical characteristic, a situation contrary to the characteristic is extracted as a risk (Fig.3, Step 3). This is done manually by the analyst for each use case. To facilitate this task, the AI ethics model has, for each AI ethical characteristic, a text of the relevant part of the ethics guideline and an example of extraction in a specific use case.

A practice guide for conducting this impact assessment process consisting of procedures, AI ethics models, and analysis sheets is published [21]. Case studies using the impact assessment to some use cases are also published.

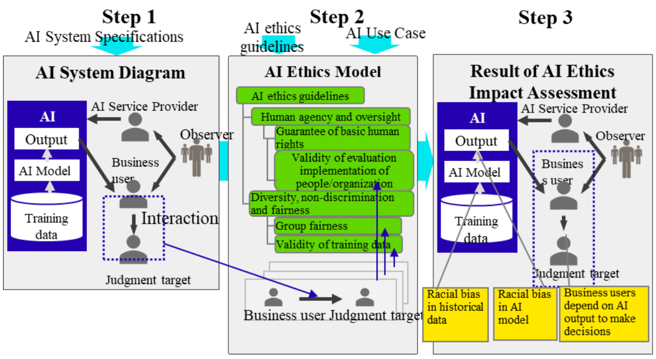


Fig. 3: Overview of the AI Ethics Impact Assessment

IV. VERIFICATION OF THE PROPOSED METHODOLOGY

We conducted initial experiments to evaluate the validity of the hypothesis presented in Section II and the ease of the impact assessment procedure. Eight subjects participated in the experiments. They all have knowledge of some AI ethical issues. The subjects were presented with nine application examples of the AI Ethics Impact Assessment and asked to answer the questions in Experiment 1 and Experiment 2, which will be described later. Application examples are the results of AI Ethics Impact Assessment conducted by an analyst on nine selected ethical issue cases registered in the AI Incident database [20]. Table I presents the details of these use cases. The column "ID" in the table indicates the index of the AI incident database.

After preparing in advance, subjects read the explanatory text on the objectives and usage scene of the AI service and the configuration of the AI system for each use case, and answer the following questions.

TABLE I. ETHICAL ISSUE CASES USED IN THE EXPERIMENTS

No	ID	Name	AI Task	Ethical issues
1	6	Chatbot	Text generation	Chatbot replies in a discriminatory chat
2	9	Teacher evaluation	Classification	The teachers' union filed a lawsuit claiming that the AI's assessment was unwarranted
3	11	Recidivism risk prediction	Classification	Black people are more likely than white people to be falsely predicted by AI as having a higher risk of recidivism
4	16	Photo tagging	Classification	Photos posted on social networks are racially tagged
5	24	Manufac-turing robot	Image recognition, environment sensing	The robot could not recognize the approaching worker and caused a contact accident
6	36	Traffic violator detection	Classification	AI wrongly detected an irrelevant person
7	37	Recruit-ment AI	Classification	Discrimination against women in screening results
8	74	Facial recognition by police	Classification	Irrelevant citizens are surely restored.
9	95	Video interview screening	Classification	Gender and race bias in heating results from video interviews

A. *Preliminary prepaction*

Subjects read instructions on how to define AI system components and stakeholders and how to view system diagrams, which are covered in the AI Ethics Impact Assessment.

B. *Experiment 1: Validity of our hypothesis*

Objectives: For each risk presented in the analysis results, to examine the validity of the interactions associated with that risk and test the validity of the hypothesis.

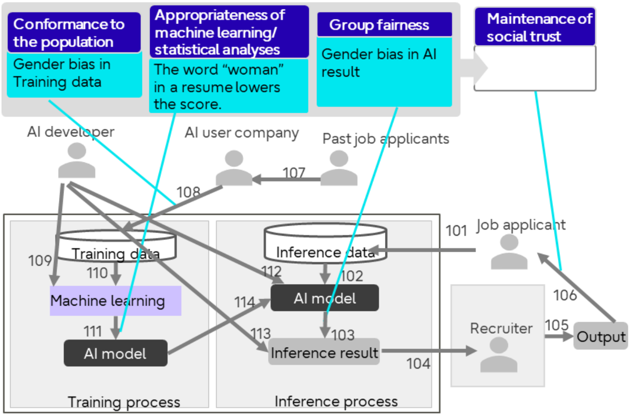
Questions: Subjects answer the following questions for each of the risks listed in the system diagram of the analysis results. In the example shown in Fig. 4, three ethical risks are described, each of which is to be answered.

- Q1-1: Are the interactions linking the risk valid? (Valid/Valid but there are other relevant interactions/Not valid and there are other relevant interactions/No relevant interactions)

- Q1-2: If you select “Valid, but there are other relevant interactions”, provide the ID of any other interactions that may apply (free format).

- Q1-3: Reasons for selecting the response from Q1-2 (free format)

- Q1-4: Are there any interactions that are not shown in the system diagram that are considered to pose risks? (free format)



Description of "Maintenance of social trust"
Does the AI system potentially negatively discriminate against people on the basis of any of the following grounds (nonexhaustively): sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation?

Fig. 4. An result of application example of recruitment AI case

C. *Experiment 2: Ease of risk extraction*

Objectives: To examine the ease of risk extraction from the AI ethical characteristics and corresponding guideline text linked to interactions.

Questions: In the system diagram shown in Experiment 1, one of the ethical risks is blank. Q2-1 asks to fill in this blank from the explanatory text of the correspond guideline.

- Q2-1: Based on the text of the guideline, assume the concrete risks and describe them in text (free format).
- Q2-2: Is that easy to answer Q2-1? (Yes/No)

D. *Results of Experiment1*

For each use case, as an indicator of the validity of the association between risk and interaction, we defined the scores for each option in Q1-1 as follows:

- (Valid) = (Total number of response with "Valid") / (Number of risks) * (Number of subjects)
- (Valid but other interactions) = (Total number of response with "Valid, but there are other relevant interactions") / (Number of risks) * (Number of subjects)
- (Not Valid and other interactions) = (Total number of response with "Not valid, and there are other relevant interactions") / (Number of risks) * (Number of subjects)
- (No interactions) = (Total number of response with "No interactions") / (Number of risks) * (Number of subjects)

Fig. 5 shows these scores for each use cases. The combined scores of "Valid" and "Valid but other interactions" exceeded 0.7. We consider this result is acceptable as a validity of the association between risk and interaction.

E. *Results of Experiment 2*

We defined the following score on Q2-2:

- Q2-2: (Ease of risk assumption) = (Number of responses with "Yes")/(Number of responses)

Fig. 6 shows "Ease of risk assumption" for each use case. Chatbots, recruitment AI, recidivism risk prediction, facial recognition by police, and photo tagging scored over 0.8. In these use cases, subjects could extract the risk "AI makes discriminatory decision" from the explanatory text shown in Fig. 4. On the other hand, teacher evaluation and manufacturing robot cases involve the risks of AI evaluating people or harming them. It is inferred that these were difficult to assume risks from descriptions of guidelines.

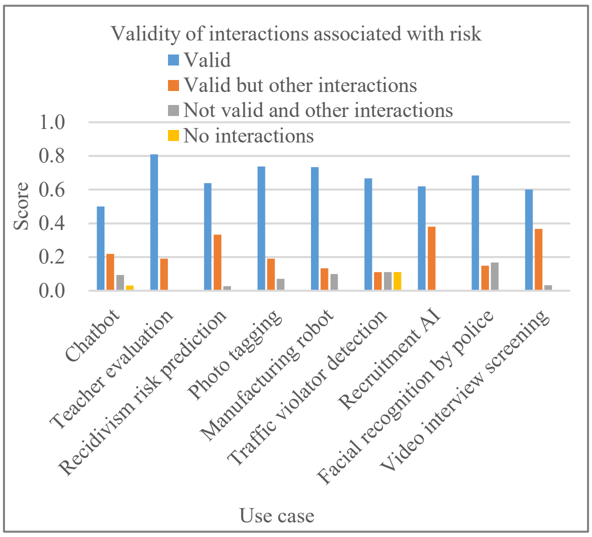


Fig. 5. Results of Experiment 1

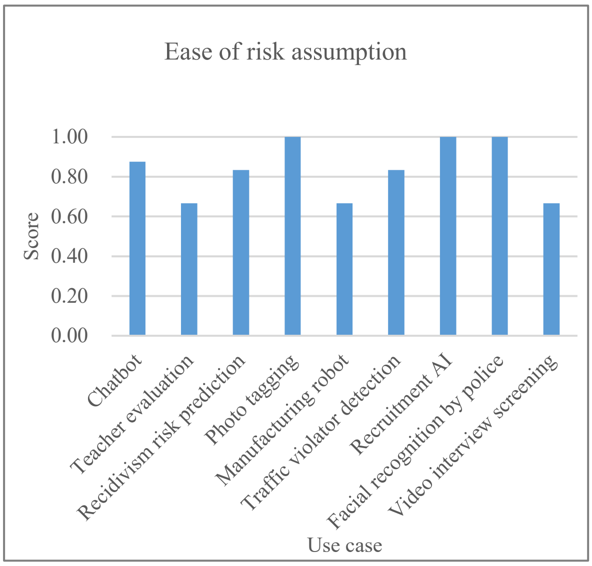


Fig. 6. Results of Experiment 2

F. *Effectiveness of associating interactions with risks in results of recruitment AI case*

The results of the responses to the recruitment AI shown in Fig.4 are discussed. the combined score of "Valid" and "Valid but other interactions" in Q1-1 was 1.0.

With regard to the risk of "Gender bias in AI result" associated with the interaction from AI model to inference result (interaction ID is103 in Fig.4, the same as hereinafter), multiple subjects who selected "Valid but other interactions" responded to Q1-2 regarding which other interactions were associated with the risk. First, the interactions from recruiter to inference result (104), from recruiter to output (105), and from output to job applicant (106) are answered that they can be associated with the risk "Gender bias in AI result". In addition, when the output from recruiter is used as retraining data, the interactions from AI user company to training data (108), from training data to machine learning (110), and from machine learning to AI model (111) were also associated with the risk. These responses suggest that associating interactions with risks on a system diagram makes it easier to comprehensively understand the effects of ethical risks.

G. *Summary of the experiments*

The experiments confirmed that the validity of the hypothesis was at an acceptable. We also confirmed that the task of assuming risks according to ethical guidelines is relatively easy for AI to make discriminatory judgments. However, in other type or ethical risks, Functions to support the work of assuming risks from the text of the guideline are necessary. It is considered to be an effective function to show risks extracted in past cases as examples.

The reliability of the response may not be sufficient enough because there were few subjects and both had knowledge of AI ethical problem cases. In the future, experiments with more diverse subjects will be necessary.

V. Summary and Future Work

This paper proposes an AI ethics impact assessment methodology that comprehensively extracts potential ethical risks in AI systems in accordance with AI ethics guidelines. Based on an analysis of past ethical issues, we hypothesized that ethical risks are associated with interactions between AI systems and stakeholders. Based on this hypothesis, we developed an AI ethics model that embodies ethical guidelines and associates it with interactions using requirement engineering approach, and developed an impact assessment procedures. As initial experiments to verify the proposed methodology, a questionnaire survey of subjects was conducted. It was confirmed that our hypothesis may be acceptable and it was possible to assume ethical risks from the explanatory texts of the ethics guidelines corresponding to the interactions. It is a future task to improve this methodology by conducting verification with more diverse subjects.

Acknowledgment

We are grateful to mediaX at Stanford University for holding a valuable workshop to advance this research.

References

[1] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F, Rossi, B. Schafer, P. Valcke, and E. Vayena, "AI4 People-An Ethical framework for a good AI society: options, risks, principles, and recommendations," Minds and Machines, 28, 689 -707, 2018.

[2] The EU High-Level Expert Group on AI, "Ethics guidelines for trustworthy AI ",https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai? April, 2019.

[3] OECD, "OECD AI Principles overview ",https://oecd.ai/en/ai-principlesSay, May, 2019.

[4] IEEE, "Ethically aligned in design First Edition", https://ethicsinaction.ieee.org/wp-content/uploads/ead1e.pdf, 2019.

[5] The conference toward AI network society, "AI Utilization Guidelines ", https://www.soumu.go.jp/main_content/000658284.pdf, August, 2019.

[6] The cabinet office of Japan, "Social Principles of Human-centric AI (Draft)", https://www8.cao.go.jp/cstp/stmain/aisocialprinciples.pdf, 2019.

[7] European Commission, "Proposal for a Regulation of the European pariliament and of the council – Laying down harmonized rules on artificial intelligence (artistic intelligence act) and amending certain union legitimative acts", https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=ENG, April, 2021.

[8] Congress.gov, "H.R.3907 - Facial Recognition and Biometric Technology Moratorium Act of 2021", https://www.congress.gov/bill/117th-congress/house-bill/3907, June, 2021.

[9] T. Kamishima, "Fairness-Aware Machine Learning and Data Mining", https://www.kamishima.net/archive/faml.pdf, April, 2022.

[10] K. Kobayashi, Y. Nakao, "One-vs-One Mitigation of Intersectional Bias: A General Method to Extend Fairness-Aware Binary Classification", DiTTE 2021, pp. 43 -54, Octover, 2021.

[11] Y. Nakao, S. Stumpf, S. Ahmed, A. Naseer, L. Strappelli, "Towards Involving End-users in Interactive Human-in-the-loop AI Fairness", https://arxiv.org/abs/2204.10464, April, 2022.

[12] Government of Canada, "Algorithmic Impact Assessment tool ", https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html, April, 2022.

[13] Ada Lovelace Institute, "Algorithmic impact assessment: a case study in healthcare", https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/, February, 2022.

[14] L. Floridi, M. Holweg, M. Taddeo, J.A. Silva, J. Mökander, Y. Wen, "capAI – A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4064091, March, 2022.

[15] R. Schwartz, A. Vassilev, K. Greene, L. Perine, A. Burt, and P. Hall, "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence", NIST Special Publication 1270, March 2022.

[16] T. Matsumoto and A. Ema, "RCModel, a Risk Chain Model for Risk Reduction in AI Services", https://arxiv.org/abs/2007.0321, 2020.

[17] K. Ohashi, A. Katayama, N. Hasegawa, H. Kurihara, R. Yamamoto, J. Doerr, and D.P. Magin, "Focusing Requirements Elicitation by Using a UX Measurement Method", 2018 IEEE 26 th International Requirements Engineering Conference, October, 2018.

[18] ISO/IEC 25010: 2011, "Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models", 2011.

[19] ISO/IEC 24030: 2021, "Information technology — Artificial intelligence (AI) — Use cases", https://www.iso.org/standard/77610.html, May, 2021.

[20] S. Mcgregor, "When AI Systems Fail: Introducing the AI Incident Database", https://partnershiponai.org/aiincidentdatabase/, November, 2020.

[21] https://www.fujitsu.com/global/about/research/technology/aiethics/, April, 2022.