A Study on Functional Requirements and Inspection Items for AI System Change Management and Model Improvement on the Web Platform

Dongsoo Moon and Seongjin Ahn*

Dept. of Computer Education, Sungkyunkwan University, 25-2 Seonggyungwan-ro, Jongno-gu, Seoul, Republic of Korea
E-mail: m1d2s3@skku.edu.com; sjahn@skku.edu.com
**Corresponding Author*

Received 16 June 2024; Accepted 21 July 2024

Abstract

The rapid adoption of artificial intelligence (AI) on the web platform across multiple sectors has highlighted not only its inherent technical hurdles, such as unpredictability and lack of transparency, but also significant societal concerns. These include the misuse of AI technology, invasions of privacy, discrimination fueled by biased data, and infringements of copyright. Such challenges jeopardize the sustainable growth of AI and risk the erosion of societal trust, industry adoption and financial investment.

This analysis explores the AI system's lifecycle, emphasizing the essential continuous monitoring and the need for creating trustworthy AI technologies. It advocates for an ethically oriented development process to mitigate adverse effects and support sustainable progress. The dynamic and unpredictable nature of AI, compounded by variable data inputs and evolving distributions, requires consistent model updates and retraining to preserve the integrity of services.

Journal of Web Engineering, Vol. 23_6, 831–848.
doi: 10.13052/jwe1540-9589.2366

Addressing the ethical aspects, this paper outlines specific guidelines and evaluation criteria for AI development, proposing an adaptable feedback loop for model improvement. This method aims to detect and rectify performance declines through prompt retraining, thereby cultivating robust, ethically sound AI systems. Such systems are expected to maintain performance while ensuring user trust and adhering to data science and web technology standards. Ultimately, the study seeks to balance AI's technological advancements with societal ethics and values, ensuring its role as a positive, reliable force across different industries. This balance is crucial for harmonizing innovation with the ethical use of data and science, thereby facilitating a future where AI contributes significantly and responsibly to societal well-being.

Keywords: AI ethics, AI model improvement, AI retraining, AI feedback loop, functional requirements and inspection items for AI.

1 Introduction

1.1 Understanding of AI Systems and Their Current Spread

AI technology follows a staged cycle, as illustrated in Figure 1, which is utilized to build and organize a learning model capable of detecting correlations and patterns among features using substantial volumes of training data. Upon receiving input data from users following the deployment of AI-based services on the web platform, it analyses the input data based on a training model and provides prediction outcomes or recommendations.

AI technology, encompassing machine learning and generative AI on web platforms, holds the potential to drive innovation across a broad spectrum of industries, including construction, fashion, graphic design, and more. It can streamline business processes, enhance creativity, and generate

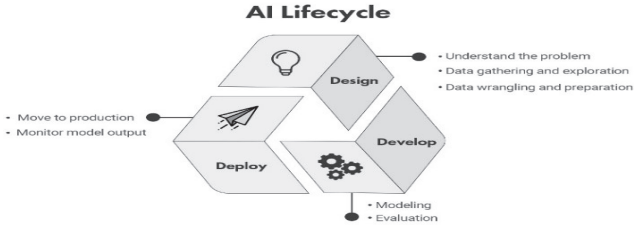

Figure 1 AI guide for government [1].

novel solutions to intricate challenges. Despite its numerous advantages, this technology involves ethical concerns in fields such as finance and healthcare, including racial and gender discrimination, violation of personal information and intellectual property rights, and generation of incorrect information, leading to distorted perceptions and conflicts [2].

As AI technology progresses, deciphering AI mechanisms becomes increasingly challenging. Hence, diverse stakeholders should engage in continuous monitoring of AI systems on web platforms to mitigate the adverse social effects of AI technology.

1.2 Social Issues on the Risks of AI

Geoffrey Hinton, a prominent figure in the AI field, has recently quit Google, warning against the risks of AI. He mentioned that "bad actors" may exploit AI for malicious purposes, emphasizing the needs to "remain committed to a responsible approach to AI" and "continually learn to understand emerging risks while also innovating boldly." In her book Weapons of Math Destruction, O'Neil argues that AI and ML algorithms have the potential to unintentionally perpetuate existing prejudices in the vast amount of data they are trained on [4]. In her book Automating Inequality, Eubanks analyzed how the introduction of technology can infringe on individuals' privacy and aggravate opportunity disparities [5]. Accordingly, she asserts that the rapid spread of AL and ML technologies would worsen existing inequalities and widen the digital divide between those with access to advanced tools and those without access. In their report, Discriminating Systems: Gender, Race and Power in AI, West, Whittaker, and Crawford investigated AI approaches to strengthen existing inequalities, social restrictions, and discrimination [6]. They also analyzed the discriminative effects of AI technology on gender and race, focusing on algorithmic biases, data distortion, and poser structures related to technological developments [6].

1.3 The Necessities and Purposes of Research

AI models, including ML, which are distributed in a production environment to provide services, compare the new data with the standard data that they learn. If changes in user behavior or additional elements of practical interactions that can affect prediction outcomes are detected, this phenomenon is referred to as AI model drift. This drift serves as a significant factor, which leads to a decrease in the accuracy of an AI model [7, 8]. To achieve automation, these studies adopted incremental and continuous learning,

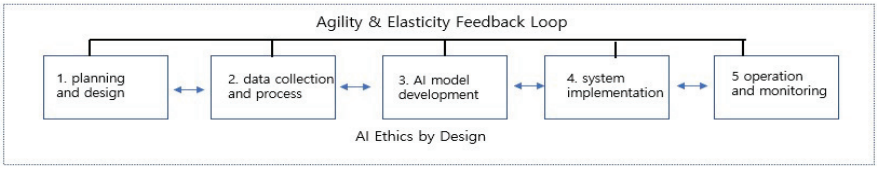

Figure 2 A feedback loop for ensuring retraining of AI systems and services.

transfer learning, retraining, and meta-learning techniques, which were used to designate flags for data drift and facilitate model retraining based on new data and all forms of manual intervention [9–11]. In addition to the issue of model drift, AI models, including those using advanced ML, may encounter variables, such as weight adjustments and labelling errors, when periodically updating their training datasets.

This study proposes functional requirements and inspection items that fulfill the necessary principles (safety, transparency, fairness, privacy protection, and responsibility) of AI ethics for the AI-SDLC.

The proposed functional requirements and inspection items were designed to solve problems that occur because of the unidirectional lifecycle stages (planning and design, data collection and processing, AI model development, system implementation, and operation and monitoring) of AI systems. Specifically, this study addresses the proposed functional requirements and inspection items based on the following classification standards: (1) essential functional requirements and inspection items for AI development, (2) functional requirements and inspection items that can be used to determine a decrease in the performance of an AI model (e.g., delivery of unintended results and occurrence of issues on bias and discrimination) in the AI system operation and monitoring stage, when data are received from users after completion of AI development and distribution of services, and (3) functional requirements and inspection items required for the establishment of procedures for a model retraining feedback loop to constantly provide yield high-quality services when deterioration of an AI model's performance is detected (Figure 2).

2 Theoretical Background

2.1 A Trend of Policies on AI Ethics in Response to the Use of AI

Simultaneously, the development and spread of AI technology have given rise to unexpected side effects from social and ethical perspectives, requiring

discussion on the implementation of trustworthy AI. To implement trustworthy AI, technical, social, and ethical aspects should be considered in the discussions. Governance can also be established to reach social consensus [12]. AI systems should ultimately protect the dignity and privacy of users and provide them with substantial benefits [13].

As the application range of AI (AI) and its influence on daily lives gradually increase, the development of trustworthy AI has emerged as a crucial task worldwide. The more AI advances, the more challenging it becomes to understand its operational principles and mechanisms. Moreover, the growing use of this technology increases the volume of accumulated data, generating issues of data pollution or bias and increasing the likelihood of errors in AI systems. The importance of establishing trustworthy AI has become more prominent, particularly because of the extensive use of AI in fields directly linked to human life and public safety [14].

In South Korea, the Ministry of Science and ICT and Telecommunications Technology Association (TTA) published 2022 Guidelines on the Development of trustworthy AI [15]. However, these guidelines are limited in that they do not present alternative solutions to ensure trustworthiness when the AI technology outputs unexpected results.

Similar cases have been previously reported. In explaining the proof-of-concept of the production gap as a project or solution to distribute ML projects in the real world, Ng pointed out that systems that operate properly in the development stage may not operate properly in the field [16].

2.2 Main Issues Regarding AI Ethics

Numerous ethical principles and guidelines have been proposed over time to serve as ethical norms that AI should follow to perform analyses or derive results. However, these circumstances have sparked growing concerns regarding the level of strict adherence to these principles and guidelines during the AI development and application stages, as well as their effectiveness in the real world [13, 17].

With the rapid supply and application of AI, various ethical issues (e.g., discrimination, fairness, violation and abuse of personal information protection, and hindrance to public interests for corporate or user interests), and relevant negative social impacts have emerged. To overcome these challenges, solutions should be developed based on the expertise and insights of diverse professionals such as philosophers, scientists, engineers, policymakers, lawyers, and social workers [18].

Moreover, as human society increasingly depends on AI, countries should reach a social consensus to prevent ethical issues in AI and systematically manage these issues based on case exchanges among governments, organizations, and companies. Concerning AI design and operation, the current ethical guidelines are too superficial and ambiguous to exert a practical influence on human decision-making [19]. In addition, there was a case in which individuals with no criminal association, particularly social minorities (e.g., people of color, homeless people, and impoverished groups), were arrested unjustly when AI was used to identify crime suspects [20]. AI also presented highly disadvantageous predictions for Black individuals in the process of predicting recidivism based on sentences and parole data. These cases raised fundamental scepticism regarding the involvement of AI in the judicial system. Consequently, people have realized that AI is likely to make unfair judgements because of its biases [21].

Moreover, the bias of AI is not an isolated incident; rather, it stems from structured factors in the AI learning process. Accordingly, extensive research has been conducted to analyze the causes of bias in AI and develop solutions to these problems [22]. AI bias is present throughout the lifecycle of AI learning [23]. AI can be applied for various purposes, ranging from searching for images on individual websites or blogs on the Internet to corporate staff management [24].

Under these circumstances, humanity is faced with the challenge of establishing clear ethical guidelines for determining the legitimacy of data selection, manipulation, and analysis in the process of using big data [25]. Moreover, discussions should be conducted to address the dual nature of big data usage, which has the potential to infringe upon personal privacy [26]. The vision of an information-oriented society in the future comprises networks that encompass everything, along with systems that can record and supervise the ideas, intentions, behaviors, and movements of its members within the network [27].

Park argues that individuals will highly depend on technical tools for their ideas, emotions, learning, and communication in the future society [28]. In this regard, he argues that ethical interest in the use of technology in future society should focus on not only privacy protection but also innovative life changes (e.g., metaverse) brought about by the new environment highly dependent on technology [29]. In this regard, we should prepare comprehensive ethical education, which encompasses issues on changes in human life and values beyond those on data and privacy protection, to effectively overcome challenges on AI ethics.

Table 1 Establishment of goals for achieving ethical and safe AI values

Goals	Contents
Ethically acceptable system design (ethically permissible)	Ensuring that AI projects have a positive impact on key stakeholders and communities, to the extent that they are ethically acceptable
Fairness and non-discriminatory system design (fair and non-discrimination)	Recognizing and working to minimize negative impacts, designing systems to minimize discriminately biased benefits to specific individuals or groups
Designing a trust-building (worthy and public trust)	Gain public confidence (solid, secure, reliable, secure) in the resulting final outcome
Designing a justification system (justifiable)	Ensure ethical legitimacy in AI system design and implementation based on transparent and interpretable evidence

2.3 Necessity of AI Ethics, Functional Requirements, and Inspection Items for AI Development

The UK, a leading country in AI ethics, strives to achieve fair AI use by developing guidelines for safe AI use in response to its reported side effects. For example, the guidelines on AI ethics for user protection offer specific measures for detecting and preventing the adverse impacts and potential risks of AI use [30].

The Alan Turing Institute defines AI ethics as a set of standards that determine the moral and ethical conduct allowed during the development and application of AI systems (Table 1) [31]. AI systems should be developed to minimize their negative effects, such as incorrect algorithm design and data bias, and to meet the social benefits related to safety and ethical aspects without focusing on maximizing efficiency and productivity [30]. Thus, they consider AI ethics as a collection of moral and ethical values, principles, and techniques that are socially accepted in AI system design, highlighting the need to minimize the damage to individuals and society caused by AI.

3 Materials and Method

3.1 Proposal of an AI System Development and Retraining Model and its Elements

In this study, we propose functional requirements and inspection items for an AI-software development life cycle (AI-SDLC) based on a feedback loop for the planning and design of AI and its retraining in the AI operation

Table 2 Functional requirements for the development of AI systems and services

Classification	Requested Item	Functional Requirements
(Step 1) Planning and design	(Requirement 01) Planning and carrying out risk management for AI systems	(01-1) Did you analyze the risk factors that may appear over the life cycle of the AI system? Attach Appendix A for detailed functional requirements

and monitoring stages after the distribution of AI systems and services. Subsequently, we conducted two rounds of focus group interviews (FGIs) with a panel of nine experts to collect and classify their opinions on the model research. Through these processes, we derive the functional requirements and inspection items for the AI-SDLC. The experts who participated in the FGIs included an expert in AI ethics, four AI instructors, an AI designer, and three AI developers.

We designated the functional requirements and inspection items to be considered depending on each life cycle stage of the AI system development by referring to the 2022 Guidelines on the Development of Trustworthy AI developed by the Ministry of Science, ICT, and TTA in South Korea [32].

3.1.1 Proposal of Functional Requirements for Development of AI Systems and Services

Table 2 lists the functional requirements for ethical AI system development depending on the following stages of the AI-SDLC: planning and design, data collection and processing, model development, system implementation, and operation and monitoring. The details of the functional requirements for each stage are provided in the Appendix.

3.1.2 Proposal of Functional Requirements and Inspection Items for an AI Retraining Feedback Loop for AI System Model Improvement

After analyzing the functional requirements and inspection items for model improvement (retraining) described in Table 3, and confirming the necessity of model improvement (retraining) through discussions with various stakeholders, we established a feedback loop for AI model improvement, as shown in Table 5. An abbreviated version of the developed feedback loop is listed in the main text, and the corresponding details are presented in the Appendix.

Table 3 Functional requirements and check items of feedback loop for AI system model improvement (retraining) example

Classification	Feature Requirements	Inspection Item
Operation and monitoring	(16-1) Are monitoring and retraining methods in place to maintain the performance of the AI system?	(Regular retraining) Regular retraining should be performed through reorganization of organizations and tasks within the institution, such as monthly, quarterly, first and second half, and once a year. Attach Appendix C for detailed functional requirements

3.2 Analysis Strategy

We established functional requirements and inspection items for the AI-SDLC from an ethical perspective and surveyed those currently working in AI-related fields at the time of the research to examine the validity of these standards. The purpose of this survey was to determine whether the respondents consistently addressed the necessity of establishing functional requirements and inspection items. Cronbach's alpha was used to confirm the consistency of the responses to the inspection items presented for each requirement. In particular, Cronbach's alpha was designed to measure the reliability of the survey items during the survey phase. This methodology was used to determine whether the inspection items obtained reliability from experts working in AI-related fields. Particularly, we focused on the following research questions:

- A. What are the necessary functional requirements (see Table 2) for AI system development ranging from stage 1 (planning and design) to stage 5 (operation and monitoring) from an ethical perspective?
- B. What are the functional requirements and inspection items (see Table 3) that can be used to determine a decrease in the performance of an AI model in stage 5 (operation and monitoring) after the distribution of an AI system to users?
- C. What are the functional requirements and inspection items (refer to Table 5) for an AI model retraining feedback loop that encompasses the stages of data collection and processing, model development, and system implementation.

3.3 Survey Methods and Targets

Survey questionnaire items were constructed to verify inspection items for the five stages (planning and design, data collection and processing, AI model development, system implementation, and operation and monitoring) of the AI-SDLC, which were derived and organized using FGIs.

The questionnaire was divided into three sections based on the research questions. In the first section, respondents were instructed to identify the necessary functional requirements for the initial AI system development, spanning from stages 1 to 5, which was the main focus of [Research Question A]. Specifically, the items in this section were included to determine the functional requirements that satisfy the principles of AI ethics for initial AI development. In the second section, respondents were instructed to identify the functional requirements and inspection items that could be used to determine a decrease in the performance of an AI model during stage 5 (operation and monitoring), which was the main focus of [Research Question B]. In the third section, the respondents were instructed to identify the necessary functional requirements and inspection items for a feedback loop encompassing each stage of the AI-SDLC for model retraining, which were the main focus points of [Research Question C]. The survey items were measured on a 5-point scale (levels: not at all, not much, moderately, rather much, or very much). The survey questionnaire was designed to investigate the consistency and reliability of the respondents' responses regarding the necessity of functional requirements and inspection items for AI system development based on the principles of AI ethics.

3.4 Demographic Characteristics and Descriptive Statistics Analysis

3.4.1 Demographic Characteristics

27 participants took part in the survey, including 17 males (63%) and 10 females (37%), participated in the survey. Table 4 presents the demographic characteristics of the participants.

3.4.2 Analysis of Functional Requirements to be Considered for Each Stage of AI Development

As shown in Table 2, the AI-SDLC is classified into stages 1 (planning and design), 2 (data collection and processing), 3 (AI model development), 4 (system implementation), and 5 (operation and monitoring). We then established the necessary functional requirements for each stage from an ethical

Table 4 Demographic characteristics

Division		Frequency (Persons)	Percent (%)
Gender	Male	17	63
	Female	10	37
	Total	27	100
Education	Doctor's degree	3	11
	Master's degree	8	30
	Bachelor's degree	16	59
Job	Professor	2	7.7
	Teacher	7	27
	Stakeholder	10	27
	AI experience	7	38
	Experience	14 (4, 18) median (IQR)	

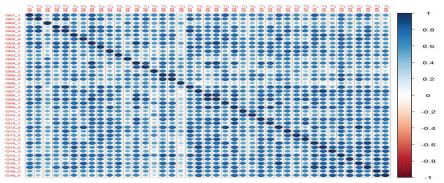

Figure 3 A correlation plot for a correlation analysis related to [Research Question A].

perspective and analyzed the reliability of these requirements based on survey responses (Figures 3–5).

3.4.3 Analysis of Functional Requirements and Inspection Items, Which Can be Used to Determine a Decrease in the Performance of an AI Model During the Operation and Monitoring Stage After Completion of AI Development and Distribution of the Developed AI System

We analyzed the reliability of the functional requirements and inspection items, as presented in Table 3, which can be used to determine the necessity of retraining (adjustment management and model improvement) of an AI system during the operation and monitoring stages after its distribution to users.

3.5 Reliability Analysis

With the rapid advancement of AI technology, concerns regarding AI ethics are expected to rise. We examined the main points of discussion on AI

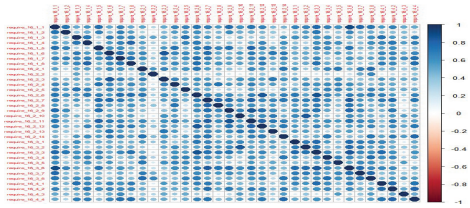

Figure 4 A correlation plot for a correlation analysis related to [Research Question B].

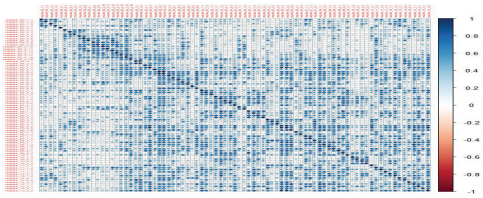

Figure 5 A correlation plot for a correlation analysis related to [Research Question C].

ethics by reviewing South Korean and international guidelines, policies, charters, research papers, service agreements, and other forms of literature related to AI ethics, development, and research. Based on the analytical results, we established functional requirements and inspection items that can satisfy the three necessary perspectives for AI, i.e., development, operation, and monitoring, and a feedback loop for AI retraining. Considering the numerous inspection items developed in this study, descriptive statistics on survey responses according to inspection items are described in detail in the Appendix. In terms of the descriptive statistics, the mean values were clustered at approximately 4. The standard deviation values range from 0.5 to 0.7, indicating a significantly stable distribution. These results indirectly suggest that the respondents found the inspection items to be reliable.

Specifically, we derived the functional requirements for each stage of the AI-SDLC based on the principles of AI ethics. Subsequently, we selected functional requirements and inspection items that could be used to determine the decrease in the performance of an AI system to identify the necessity of retraining it. Furthermore, we designated the functional requirements and inspection items for a feedback loop for AI model improvement, which can solve the problem of decrease in the performance of the AI system when the necessity for AI retraining is determined. Table 5 presents the results of the reliability analysis based on the established functional requirements and inspection items.

Table 5 Reliability analysis results

Classification	Feature Requirements	Number of Check Items	Cronbach's α
Operation and monitoring	(16-1) Are monitoring and retraining methods in place to maintain the performance of the AI system?	8	0.895
	Attach Appendix D	Attach Appendix D for detailed inspection items	

We conducted a reliability analysis of the survey results on the functional requirements and inspection items required for a feedback loop encompassing each stage of the AI-SDLC for AI model retraining and improvement..

4 Conclusions

This study delineates the essential functional requirements and inspection protocols for AI system development, emphasizing an ethical framework. We propose a feedback loop for swiftly and adaptively enhancing AI systems, particularly critical after identifying retraining needs for AI models during their operation and monitoring phases, a process increasingly implemented on web platforms.

AI systems evolve through a complex lifecycle, encompassing planning, design, data collection, processing, AI model creation, deployment, and continuous operation and monitoring. Our focus is on pinpointing the ethical functional requirements and inspection criteria vital for the development of AI technologies, aiming to mitigate adverse effects and promote ongoing advancement.

These requirements and criteria are categorized as follows: (i) those essential throughout the five stages of an AI system development life cycle (AI-SDLC), (ii) those aimed at managing AI's inherent uncertainty, especially in stage 5 (operation and monitoring) after system rollout, to ensure constant human oversight of unexpected events and determine AI retraining needs, and (iii) those needed for a feedback loop from stages 2 to 5 to efficiently address performance declines in AI models.

To foster human-centric and trustworthy AI on web platforms, an urgent pivot towards AI ethics awareness is required. AI systems and services should be developed within a framework of appropriate legal, systemic, ethical, and technical guidelines, ensuring adherence to values of safety, transparency,

fairness, privacy, and accountability. Establishing actionable development guidelines will enable stakeholders from diverse sectors to participate in evaluating and anticipating the user-centric planning, design of AI, and its social consequences throughout the AI-SDLC. These measures aim to bolster the trustworthiness and marketability of AI technology, facilitating its sustainable development and fostering the creation and utilization of socially responsible AI, thereby minimizing and swiftly addressing AI-related harm. Ethically developed and deployed AI assures user safety and dignity, offering significant benefits characterized by fairness and the absence of bias or discrimination.

The ideal ethical framework for AI systems and services is shaped by the collective moral perspectives of all stakeholders involved, including designers, developers, and operators. As AI technologies increasingly permeate various fields, further research is imperative to establish performance evaluation guidelines at every AI-SDLC stage. Such endeavors are expected to contribute profoundly to the development of human-oriented AI in the context of the fourth industrial revolution's intelligent, informatized society.

References

[1] AI Center of Excellence. AI Guide for Government: A Living and Evolving Guide to the Application of AI for the U.S. Federal Government, 2023.
[2] Lu, J. Liu, A. Dong, F. Gu, F Gama, J. Zhang, G. Learning under Concept Drift: A Review. IEEE Trans Knowl Data Eng. 2019, 31, 2346–2363.
[3] Gartner Research. Infographic: AI Use Case Prism for Retail. 2021.
[4] O'Neil. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. 2016.
[5] Eubanks. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. 2018.
[6] West, S. M.; Whittaker, M.; Crawford, K. Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. 2019. Available Online: https://ainowinstitute.org/discriminatingsystems.html. (accessed on 16 June 2023).
[7] Raţiu S, Aurel R, Cătălinoiu, Zsolt Miklos I. "Self-Adaptive Mechanical Reducer With Variable Gear Ratio." Journal of Automotive Engineering. 2021, 66.

[8] Greco, Salvatore, et al. "On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness." Social indicators research 141 (2019): 61–94.
[9] Nagabandi, A. Finn, C. Levine, S. Deep Online Learning via Meta-Learning: Continual Adaptation for Model-Based RL. In Proceedings of the ICLR 2019, New Orleans, USA, 6–9 May 2019; Available online: https://iclr.cc/Conferences/2019/Schedule?showEvent=1078 (accessed on 23 June 2023).
[10] Hospedales T, Antoniou A, Micaelli P, Storkey A. Meta-Learning in Neural Networks: A Survey. IEEE Trans. Pattern Anal. Mach. 2022, 44, 5149–5169.
[11] Mai Z, Li R, Jeong J, Quispe D, Kim HW, Sanner S. Online continual learning in image classification: An empirical survey. Neurocomputing. 2022, 469, 28–51.
[12] Arnold Z, Toner H. AI Accidents: An Emerging Threat. Center for Security and Emerging Technology. 2021. Available online: https://doi.org/10.51593/20200072 (accessed on 3 July 2023).
[13] Lepri B, Oliver N, Pentland A. Ethical machines: The human-centric use of AI. iScience. 2021, 24, 3.
[14] Brundage M, Avin S, Clark J, Toner H, Eckersley P, Garfinkel B, Dafoe A, Scharre P, Zeitzoff T, Filar B, Anderson H, Roff H, Allen GC, Steinhardt J, Flynn C, Héigeartaigh S Ó, Beard S, Belfield H, Farquhar S, et al. The Malicious Use of AI: Forecasting, Prevention, and Mitigation. Apollo – University of Cambridge Repository. 2018.
[15] Ministry of Science and ICT · Korea Information and Communication Technology Association, (2022), 2022 Trustworthy AI Development Guide (draft).
[16] Ng A. Healthcare's AI Future: A Conversation with Fei-Fei Li & Andrew Ng. Stanford University Human-Centered AI. 2021.
[17] John-Mathews JM, Cardon D, Balagué C. From Reality to World. A Critical Perspective on AI Fairness. J Bus Ethics. 2022, 178, 945–959.
[18] McLennan S, Lee MM, Fiske A, Celi LA. AI Ethics Is Not a Panacea. Am J Bioeth. 2020, 20, 20–22.
[19] Borenstein J, Howard A. Emerging challenges in AI and the need for AI ethics education. AI Ethics. 2021, 1, 61–65.
[20] Hagendorff, T. The Ethics of AI Ethics: An Evaluation of Guidelines. Minds & Machines. 2020, 30, 99–120.

[21] Angwin J, Larson J, Mattu S, Kirchner L. Machine Bias. In Ethics of Data and Analytics: Concepts and Cases, 1st ed.; Kirsten Martin, Ed.; Auerbach Publications: New York, USA, 2022; pp. 254–264.
[22] Cheng Z, Zhu X, Gong S. Surveillance Face Recognition Challenge. Technical Report 2018. Available online: https://qmul-survface.github.io/ (accessed on 5 July 2023).
[23] Dilmegani C. Bias in AI: What It Is, Types and Examples of Bias, and Tools to Fix It. Science and Technology. 2022, 52, 1–3.
[24] Leslie D, Mazumder A, Peppin A, Wolters MK, Hagerty A. Does "AI" Stand for Augmenting Inequality in the Era of Covid-19 Healthcare? BMJ. 2021, 372.
[25] Chang K. AI in Personnel Management: The Development of APM Model. The Bottom Line. 2020, 33, 377–388.
[26] Mitov A. Ethical Use of AI Through the Utilitarianism Perspective. Bachelor's thesis, University of Twente, Netherland. 2021.
[27] Markham AN, Tiidenberg K, Herman A. Ethics as Methods: Doing Ethics in the Era of Big Data Research – Introduction. Social Media+Society. 2018, 4.
[28] Park HY. New Arguments in AI Ethics and Related Essential Considerations. The Jour. Yeolin Edu. 2022, 30, 173–201.
[29] Feldstein S. The Global Expansion of AI Surveillance. Carnegie Endowment for International Peace, 2019. Available online: https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847 (accessed on 23 July 2023).
[30] Leslie, D. Understanding AI ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. 2019. Available online: https://doi.org/10.5281/zenodo.3240529 (accessed on 29 July 2023).
[31] Alan Turing Institute. 2019.
[32] ISO/IEC TR24027:2021, Information technology – AI (AI) – Bias in AI systems and AI aided decision making, 2021.

Biographies

Dongsoo Moon received his bachelor's degree in computer science from Seoul National University of Science and Technology in 2004 and his master's degree in electronic computer education from Hankuk University of Foreign Studies in 2008. He is a Ph.D. candidate in the Department of Computer Education at Sungkyunkwan University in 2023. His primary research

areas include artificial intelligence ethics, AI system design and development, and AI and SW education.

Seongjin Ahn received his Bachelor's, Master's, and doctoral degrees in information engineering from Sungkyunkwan University. Currently, he is a professor in the Department of Computer Education at Sungkyunkwan University. His primary research areas include artificial intelligence ethics, AI and SW education, and network management.