ISSN 2435-3345

Northeast Asian Studies

Robotics & AI Ethics

Vol. 8

1. **Uprising of ChatGPT and Ethical Problems**
/ Jungbae Bang, Gyunyeol Park

2. **Suggestions for Ethical Decision-Making Model through Collaboration between Human and AI**
/ Hyunsoo Kim

3. **Soft Power in Northeast Asia, Using AI in Information Warfare**
/ Sunggu Jo

J-INSTITUTE

2023 Volume 8, 1-11 Submission: 2023/04/30, Peer review: 2023/05/30, Accepted: 2023/06/20, Published: 2023/08/30

Robotics & AI Ethics

Publisher: J-INSTITUTE
ISSN: 2435-3345

Website: j-institute.org
Editor: admin@j-institute.org

Corresponding author*
E-mail: pgy556@hanmail.net

DOI Address:
dx.doi.org/10.22471/ai.2023.8.01

Uprising of ChatGPT and Ethical Problems

Jungbae Bang[1]
Open Cyber University of Korea, Seoul, Republic of Korea
Gyunyeol Park[2*]
Gyeongsang National University, Jinju, Republic of Korea

Abstract

Purpose: *This study aims to examine ethical issues following the emergence of ChatGPT, evaluate ChatGPT with a focus on moral competence, and seek ethical solutions.*

Methods: *This study tries to use various literature and various media reports from the East and West dealing with the technology, operation method, and ethical issues of ChatGPT. In the detailed analysis of moral problems following the emergence of ChatGPT, the satisfaction of each element was evaluated based on moral competence, and alternatives were presented.*

Result: *As a result of the study, the ethical issues that can be raised following the emergence of ChatGPT include the possibility of plagiarism and copyright infringement, damage to the fairness of the test, use for criminal purposes, occurrence of social stereotypes and unfair discrimination, invasion of personal privacy and organization's Security exposure, reduced critical thinking, and loss of genuine human relationships. And as a result of evaluating the ethical issues of ChatGPT centering on moral competence, it is evaluated that moral identity, moral sensitivity, and moral practice are possible to implement, but moral judgment is evaluated to have many limitations. In order to solve these ethical problems, a utilitarian approach was proposed.*

Conclusion: *The most optimal decisions and actions related to the design, development, adoption, deployment, maintenance and evolution of ChatGPT should do the most good or the least harm to society. To do this, responsible AI toolkits and frameworks must have an ethical perspective built in, allowing for a balanced view of what is right and wrong. Along with this, a multi-stakeholder approach is needed to create a good AI society.*

Keywords: ***ChatGPT, Moral Competence, Moral Identity, Moral Sensitivity, Moral Judgement***

1. Introduction

ChatGPT(Generative Pre-trained Transformer) is a chat bot introduced in November 2022 by OpenAI, an AI research and development company, and is based on a variant of GPT-3[1]. ChatGPT is a large language model(LLM), a machine learning system capable of autonomously learning from data and generating sophisticated and seemingly intelligent writing after training on massive text data sets. It is the latest in a series of models released by OpenAI, an AI company based in San Francisco, California, and others. ChatGPT generated excitement and controversy because it was one of the first models to converse persuasively with users in English and other languages on a variety of topics[2].

One of the key advantages of these large-scale language models is their ability to understand the context of a given prompt and generate an appropriate response. This is a vast improvement over previous language models, which were often unable to interpret the meaning and intent behind a given piece of text. Another important aspect is the ability to generate high-quality text that is difficult to distinguish from human writing. The ability to elicit knowledge and

answer difficult academic questions is inherent in the ability to answer questions that cannot be easily found through a web search and to provide accurate and reliable answers[3]. This surprised people with a level of ability never seen before.

Due to these groundbreaking capabilities, ChatGPT brought an explosive response, surpassing 100 million monthly activity users(MAU) within two months of its launch. It took nine months for TikTok and two and a half years for Instagram to reach 100 million MAU, which is a tremendous speed[4]. ChatGPT is already expanding its territory in various fields. It is used in various creative fields, from writing college homework, self-introduction letter for employment, and politician's speech, to composing, drawing, and app development. A judge in Colombia made headlines by revealing that he used ChatGPT to write a ruling. It is also said that ChatGPT provided a significant part of the background knowledge needed to write the most refined and accurate judgment[5].

This increased level of personalization can lead to improved customer service and education as ChatGPT can be trained to better understand and respond to each user's specific needs and preferences. Additionally, by leveraging the vast amount of data generated from ChatGPT's interactions, developers can create language models tailored to each user's specific needs and preferences, providing a more personalized and engaging experience[6]. From these integrations with other AI technologies, to increased personalization and customizability, to continued advancements in language model performance, there are many exciting opportunities for ChatGPT technology to improve our lives in meaningful and positive ways.

Nevertheless, ChatGPT, like a 'double-edged sword', contains benefits and risks that can cause harm to humans at the same time[7]. However, previous studies have tended to make partial analyzes on the ethical issues of ChatGTP, mainly by research field. Therefore, this study aims to comprehensively examine the problems of ChatGPT from an ethical point of view. In other words, the purpose of this study is to predict what kind of ethical problems will arise due to the emergence of ChatGPT in the future, to on and to think about solutions. To this end, this study focuses on ethical issues that may arise in the process of developing and using ChatGPT, not ChatGPT itself. And after categorizing and extracting ethical problems based on various previous studies and media data related to this, the level of problems is evaluated based on the elements of moral competence, and alternatives are proposed to solve them.

2. How is ChatGPT Different from Existing AI?

AI is an umbrella term and broad field that refers to the creation of intelligent systems capable of performing tasks that normally require human intelligence, such as learning, problem solving, and decision making[8]. Artificial intelligence is the intelligence exhibited by artificial entities to solve complex problems, and these systems are generally considered computers or machines. Intelligence is the ability to think, imagine, remember, understand, recognize patterns, make choices to adapt to change, and learn from experience. Artificial intelligence that makes computers behave more like humans and in much less time than humans. That is why it is called artificial intelligence[9]. AI algorithms are used in various fields because they can process vast amounts of data, identify patterns, and make predictions beyond the capabilities of human thinking[10].

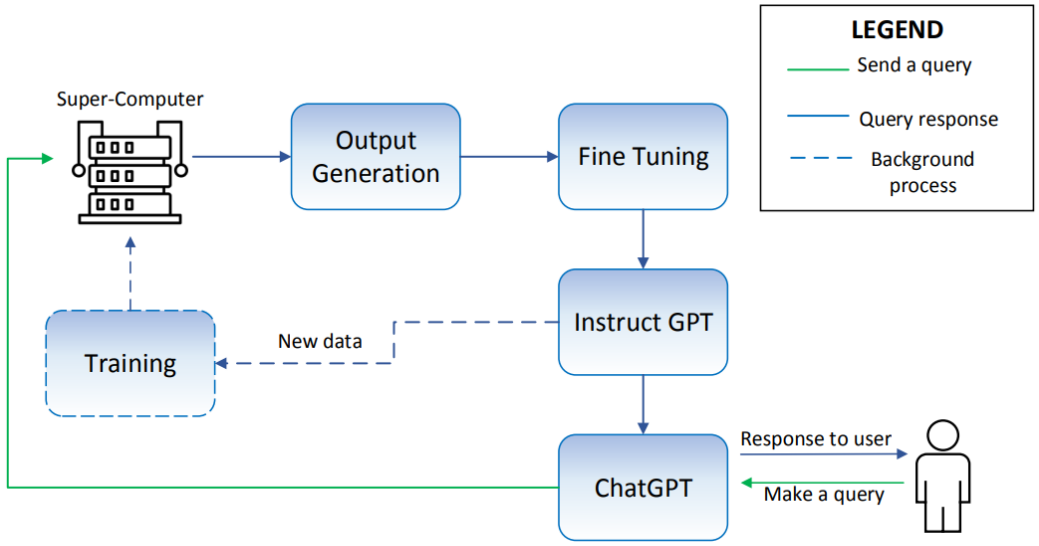
ChatGPT is a powerful conversational AI system using deep learning[11]. It is one of the AI-based conversational programs that can generate human-like responses[10]. Until now, conversational AI has often misunderstood what people say. Humans do not have perfect command of grammar and spelling, and due to the nature of language, the difference in interpretation is very wide, so there are many cases in which a person gives a written answer or cannot answer at all, so its utilization is extremely limited except for short answers[12].

However, ChatGPT did a lot of learning in the language part, applied the best AI, memorized the context of the speech and previous conversations, and reached the level of understanding even if it was spoken roughly like a conversation. It has the ability to not only generate natural language responses from questions posed in natural language, but also predict the direction of the current conversation[13]. In particular, it has the ability to provide personalized dialogue and language responses based on the different conversation styles of individual users[14]. ChatGPT can learn and analyze massive amounts of linguistic data from various sources and generate output in a human-like way. Unlike conventional AI, which simply analyzes objects and identifies patterns, ChatGPT can create new and unique objects and effects, making it a powerful generative AI[15].

The working process of ChatGPT can be divided into two types, query and response, as shown in <Figure 1>[16]. The device behind ChatGPT is an artificial intelligence supercomputer. The computer is trained on a large data set with numerous parameters. This supercomputer is trained unsupervised to identify patterns in input data by determining the statistical structure within the data. In general, users can write queries to ChatGPT. This query is sent directly to the supercomputer. Queries are now processed on supercomputers. The output generation circuitry generates possible outputs and then fine-tunes the output data. Then instruct ChatGPT to respond. Finally, ChatGPT, a conversational interface, interacts with humans by providing human-like responses.

Helberger(2023) argues that ChatGPT, a generative AI system, differs from 'traditional' AI systems in at least two important ways: dynamic context and scale of use[17]. Generative AI systems are not built for specific situations or conditions of use, and their openness and ease of control allow them to be used on an unprecedented scale. The output of generative AI systems is virtually indistinguishable from human-generated content, as they are trained using almost anything available on the web[18]. It can also be interpreted as media(text, audio, video) by people with common communication skills, significantly lowering the threshold for who can be a user. And they can be used for a certain amount of versatility because of the sheer volume of data extraction that goes into training. ChatGPT alone contains 300 billion words, spanning all kinds of content available on the Internet, from personal data to policy documents, news reports, literary texts and art.

Figure 1. Working process of ChatGPT[16].



3. Ethical Problems and Evaluation with Current ChatGPT

3.1. Moral competence as a criterion of analysis

Ethical issues of artificial intelligence can be said to be related to the morality of the person who develops and uses the system rather than the morality of the system itself [19]. The answer of artificial intelligence is ultimately learned by algorithms developed by humans and data given by humans, so its source is humans. Therefore, the ethical issues of ChatGPT, a powerful conversational artificial intelligence, should be analyzed based on the moral competence of people involved in the development and use of ChatGPT.

Moral competence is the ability of a moral agent to demonstrate desirable thoughts, emotions, and behaviors in oneself or in a community[20]. The first to use the term, L. Kohlberg(1964) defines moral competence as “the ability to make moral judgments based on internal moral principles and the ability to act in accordance with those judgments”[21]. Moral psychologist J. R. Rest(1986) presented the concept of four components of morality[22]. They are moral sensitivity, moral judgment, moral motivation, and moral practice. K. Y. Park(2017) included the concept of moral motivation, which prioritizes the moral over the immoral, as a sub-concept of moral judgment among the four components of Rest, and moral identity, which is the perception of moral agents, and one's self[20]. A new element was added, the concept of moral responsibility, which is responsible even for bad results caused by good will. However, given that moral responsibility can be subsumed as a sub-concept of moral practice, this thesis discusses the moral competence related to ChatGPT, focusing on four components: moral identity, moral sensitivity, moral judgment, and moral practice.

Moral identity is described as the degree to which morality or being a moral person is important to a person's identity. If an individual as a subject of action recognizes himself or herself as a moral person, that person is more likely to act morally due to the desire to be consistent with their self-concept[23]. Moral sensitivity is a concept that includes the sensory perception system's acceptance of a social situation and its interpretation of the situation as to what actions are possible, who and what will be affected by each possible action, and how the parties involved will respond to the possible outcomes[24]. Moral judgment involves determining which of the possible actions is moral. The individual weighs the options and decides what the person should do in those situations. Moral judgment is the ability to rank a hierarchy of values and to make weighted judgments of various kinds. In other words, it refers to the multiplier effect of value classification judgment, type classification judgment, importance classification judgment, and sequence classification judgment. It is the ethical competency that usually plays the most role. Moral practice refers to the ability to combine the forces of the ego with the social and psychological skills needed to perform selected actions[25]. It refers to the ability to move into action after going through all the previous steps.

3.2. Ethical issues about ChatGPT

Undoubtedly, ChatGPT can greatly change human life in many aspects in the future. Given its position as a universal assistant, ChatGPT could be useful in improving production effectiveness and efficiency. It is expected. It could have a major impact on almost every industry, including education, mobile, search engines, content creation, and healthcare. Despite these numerous benefits, ChatGPT may have a negative impact on human life[26]. Therefore, we have a task to carefully consider and solve the ethical problems of ChatGPT technology. Ethical issues of ChatGPT that are expected in the future can be largely classified into six categories as follows.

First, while ChatGPT has the potential to provide many advantages in higher education assessment, ChatGPT and other AI language models similar to it may raise serious ethical questions in higher education assessment. A problem with using ChatGPT for higher education assessment is the potential for plagiarism. AI essay writing systems are designed to generate

essays based on a series of parameters or prompts. That is, students can potentially use these systems to cheat on assignments by submitting essays that are not their work[27]. As such, ChatGPT can be used to further facilitate cheating, and it can be difficult to distinguish between human and machine-generated writing. With the advent of ChatGPT, it seems that online exams or take-home exams(or assignments in this context) will no longer be able to maintain test ethics and fairness[28]. In addition, ChatGPT may generate responses that violate intellectual property rights such as copyright or patent law[29]. Therefore, universities should carefully consider the potential risks and rewards of using these tools and take steps to ensure that they are used ethically and responsibly[30].

Second, ChatGPT provides easy scripting and coding access to cybercriminals, effectively reducing the barriers to entry in this space. If a malicious actor or group has access to ChatGPT, it could be used to create fake news articles, misleading social media posts or fraudulent customer reviews. This misinformation could also be due to the information used to train the AI. This can have significant consequences such as spreading misinformation, harming individuals or organizations, or influencing public opinion or decision-making[1]. You can quickly create a persuasive email or social media phishing lure, making it difficult for people to distinguish what is legitimate. ChatGPT can also be used to create fake chatbots that can be manipulated into impersonating a person or a legitimate source, such as a bank or government agency, to access sensitive information, steal money, or commit fraud[31].

Third, if the data used to train the language model includes biased representations of a specific group of individuals, social stereotypes and unfair discrimination may occur. Due to its sophisticated AI capabilities, ChatGPT has the potential to reinforce existing forms of prejudice and discrimination, which can lead to learning[32]. Because of this, there is a concern that the model provides unfair or discriminatory predictions for that group[33]. For example, language skills that analyze resumes for recruitment or career guidance may be less likely to recommend historically discriminated groups to recruiters or more likely to offer low-paying jobs to marginalized groups. Additionally, ChatGPT can often create confusing biases such as racist and sexist remarks in its answers. Even if an engineer somehow erases all explicit racism in a bot's output, it may still provide implicit racist, sexist, or other bigot bias in its output. For example, when asked to write code to evaluate whether someone would make a good scientist based on gender and race, the bot only suggests white males[34]. Recent experiments have shown that ChatGPT, despite its innate protections, can be used at large scale, including the code required for maximum spread. It can be used to create hate speech campaigns[35].

Fourth, there are risks of data security involved in interacting with ChatGPT. This may reveal personal information(age, gender, address, contact information, hobbies, capital accounts and other personal information). Most of this personal information is exposed in the user's unconscious communication process[36]. In addition, data leakage of language models can jeopardize individual privacy and organizational security in the process of exposing the model to attacks by attackers trying to extract sensitive information from the model[37].

Fifth, another ethical problem with ChatGPT is that it has the potential to reduce users' critical thinking[36]. ChatGPT's main concern is the creation of false and believable information generated by computers rather than human decision-making. As OpenAI also acknowledges, ChatGPT sometimes produces plausible-sounding but inaccurate or nonsensical answers[38]. As such, the reliability of generative language models can be compromised by hallucinations. Hallucinations refer to the creation of false or misleading information by such models. This problem is widespread in natural language generation, and misinformation and dissemination of misinformation are common manifestations of this phenomenon[39]. This may cause ethical problems for users who rely too much on the answers without recognizing the accuracy of the answers. Therefore, guidelines to promote critical thinking when using ChatGPT in the future will be needed.

Sixth, when people start to rely on machines to communicate, genuine human relationships can be lost[40]. The ability to connect with others through conversation is a fundamental aspect of being human, and outsourcing it to machines could have harmful side effects in our society. First and foremost, ChatGPT lacks the ability to truly understand the complexities of human language and conversation. It is trained to generate words based on given input, but has no ability to truly understand the meaning behind those words. This means that any response it generates is likely to be shallow and lack depth and insight. So relying on ChatGPT for conversations can be ethically questionable.

3.3. Evaluation of ChatGPT with a focus on moral competence

Artificial intelligence offers countless opportunities to improve and enhance the capabilities of individuals and society as a whole. The use of artificial intelligence technologies offers numerous possibilities for reinventing society by fundamentally improving what humans collectively can do. Because such technologies are so powerful and have the potential to be destructive, they also carry commensurate risks. Ensuring the desired outcomes of artificial intelligence in society depends on resolving the tension between incorporating the benefits of artificial intelligence and mitigating its potential harms[41]. In this context, the value of an ethical approach to artificial intelligence technology, especially the newly emerging ChatGPT technology, is clearly evident.

Ethical evaluation of ChatGPT technology can be made based on four components of moral competence: moral identity, moral sensitivity, moral judgment, and moral practice. First, since ChatGPT is used as a technological tool, it is unlikely that the emergence of ChatGPT will cause a significant change in the moral identity of the individual, the moral subject. In addition, moral sensitivity can be overcome to some extent through training, and moral practice can be implemented if specific guidelines and standards for the use of ChatGPT are prepared. However, the rise of ChatGPT for the following reasons Rather than enhancing one's moral judgment, it is feared to weaken it.

First, ChatGPT provides moral advice despite its lack of morality. In other words, ChatGPT's advice affects the user's moral judgment. People can underestimate ChatGPT's influence and adopt arbitrary moral positions as their own. Because of this, ChatGPT is more likely to weaken rather than improve users' moral judgment[42].

Second, users do not understand the information that ChatGPT generates or judge its accuracy or relevance. This may result in regulations prohibiting its use. Nonetheless, ChatGPT technology will become commonplace before institutions have time to change their policies. So instead of relying solely on AI tools to do all the necessary work, we need to let users utilize ChatGPT in a way that encourages them to engage in analytical and critical thinking. Users should ensure that they understand how AI works, as well as its capabilities and limitations, as AI continues to evolve and become widespread[43]. This should equip users with the necessary knowledge and skills to make moral judgments about the application of AI in their future professional and personal lives[32].

Third, ChatGPT, a large-scale language model, is a statistical model, not an ethical reasoning[44]. There is no a priori reason to expect ChatGPT to use the same concepts or perform the same types of reasoning as humans, and there is little evidence that it does[45]. A limitation of ChatGPT is that it cannot understand the context or meaning of the words we generate. We can generate text based on the probability of a specific word or series of words appearing together based on the given training data. That is, they cannot provide explanations or inferences about their responses and may not always produce completely coherent or meaningful responses in the context of a conversation. It would be very dangerous to relinquish control of our responsibilities as human beings to make sound moral and ethical judgments and decisions to a poorly designed system with glaring flaws[44].

3.4. Model for solving the ethical problems of ChatGPT

Once the risks of ChatGPT are identified, ethical models need to be used to determine the path forward[46]. Utilitarianism can be considered as a model for solving the ethical problems of ChatGPT. Utilitarianism is one of the most common approaches to making ethical decisions that cause the least(or best) harm to individuals, society, and the environment by considering both the positive and negative effects of an action[47]. According to utilitarianism, the most optimal decisions and actions related to ChatGPT design, development, adoption, deployment, maintenance, and evolution should do the most good or least harm to society. To do this, responsible AI toolkits and frameworks must embed an ethical perspective so that they can have a balanced view of what is right and wrong[18].

From an AI risk management perspective, the theory provides an approach to resolving the conflict through a flexible outcome-oriented lens for establishing and testing policies at each stage of the risk management cycle. It is therefore essential for organizations to understand, manage and mitigate the risks posed by AI adoption. Ethical review and bias screening should complement periodic risk assessments, as the vast amount of data used to train algorithmic models has an evolutionary nature of high velocity, heterogeneity, and variability. For example, the risk of adopting ChatGPT in a particular situation can be assessed by the Risk Management Framework(RMF), where the impact and consequences of the risk for each stakeholder can be prioritized using a utilitarian perspective. Similarly, the contextual importance of AI adoption(in each sector of a particular application) allows AI developers, organizations planning AI deployments, and even policy makers to make realistic, actionable moral decisions that can understand and assess both opportunities and negative impacts. Therefore, it is necessary to integrate AI risk management frameworks with ethical theory perspectives to make socially responsible judgments that help ensure purposeful, prudent, rational and ethical ways to leverage generative AI models such as ChatGPT[18].

Pasquale(2020)[48] argues that “algorithms must be regulated as soon as they affect the world, and programmers must take ethical and legal responsibility for the harm caused by algorithms”. However, the incomprehensibility and complexity of machine learning have hampered attempts to regulate it, and AI lacks a consensus professional code or ethical framework[49]. For example, as educational institutions begin to proliferate consumer devices, they are unlikely to actually become gatekeepers of AI technology. As students increasingly use the parameters of traditional assessments, such as essays, to overcome, educators are already beginning to incorporate language processors such as ChatGPT into their teaching. It is imperative that educators engage with the impact of generative AI on existing delivery and assessment systems. Even if all algorithms could be made transparent and fully explainable, the sociotechnical ecosystem of production, assembly, programming, training, use, and maintenance would be too decentralized to be entirely obscured from the point of view of any one individual[50]. Therefore, incorporating and making the most of large-scale language models in the classroom requires a clear strategy within the school system, a clear teaching approach that focuses on critical thinking, and a fact-checking strategy. In particular, intensive efforts are required to inform students of the potential social prejudice, moral criticism, and dangers of AI application at the beginning of their studies to enhance users' moral judgment[32].

4. Conclusion

Ultimately, ChatGPT is a tool that can help users in their daily lives, but it cannot replace the added value that humans can bring. In fact, one of the most obvious and urgent current ethical failures that exist today is the continued overstatement and mystification of technology's capabilities[51]. It is legitimate to be enthusiastic about this new technology, but it is important

to take a step back and question how it works in order to make the most of it and maintain a critical eye[52]. In this context, the purpose of this study was to examine ethical issues following the emergence of ChatGPT and to evaluate ChatGPT with a focus on moral competence.

Ethical issues that can be raised according to the emergence of ChatGPT are classified into six categories. First, there is a possibility of plagiarism and copyright infringement, and there is a concern that the fairness of the test may be damaged. Second, it can be used for criminal purposes and used for false news, fraud, phishing lure, etc. Third, there is a possibility of social stereotypes and unfair discrimination. Fourth, there is a risk of invasion of personal privacy and exposure of organizational security. Fifth, over-reliance on ChatGPT may reduce critical thinking. Sixth, there is a concern that genuine human relationships will be lost if people start to rely on machines for conversation.

As a result of evaluating the ethical issues of ChatGPT centering on these moral competence, it is evaluated that moral identity, moral sensitivity, and moral practice are feasible. But there are many limitations to moral judgment. In other words, ChatGPT affects the moral judgment of users despite its lack of morality, users do not understand the information generated by ChatGPT or do not judge the accuracy of the information. And ChatGPT is a statistical model rather than ethical reasoning. It makes moral judgment very difficult.

In order to solve these ethical problems, a utilitarian approach was proposed. According to utilitarianism, the most optimal decisions and actions regarding ChatGPT design, development, adoption, deployment, maintenance, and evolution should do the most good or least harm to society. To do this, responsible AI toolkits and frameworks must embed an ethical perspective so that they can have a balanced view of what is right and what is wrong. Along with this, we believe that a multi-stakeholder approach[41] is necessary to create a good AI society. Intensive attention from multiple stakeholders is required so that AI can meet the needs of society and increase users' moral judgment by allowing developers, users, and rule makers to participate and jointly conduct evaluations from the beginning.

5. References

5.1. Journal articles

[1] Sebastian G. Do ChatGPT and Other AI Chatbots Pose a Cybersecurity Risk?: An Exploratory Study. *International Journal of Security and Privacy in Pervasive Computing*, 15(1), 1-11 (2023).

[2] Van Dis EA & Bollen J & Zuidema W & Van Rooij R & Bockting CL. ChatGPT: Five Priorities for Research. *Nature*, 614(7947), 224-226 (2023).

[6] Aljanabi M. ChatGPT: Future Directions and Open Possibilities. *Mesopotamian Journal of Cyber Security*, 2023, 16-17 (2023).

[7] Shen Y & Heacock L & Elias J & Hentel KD & Reig B & Shih G & Moy L. ChatGPT and Other Large Language Models are Double-edged Swords. *Radiology*, 307(2), e230163 (2023).

[8] Holzinger A & Keiblinger K & Holub P & Zatloukal K & Müller H. AI for life: Trends in Artificial Intelligence for Biotechnology. *New Biotechnology*, 74, 16-24 (2023).

[9] Strong AI. Applications of Artificial Intelligence & Associated Technologies. *Science*, 5(6), 64-67 (2016).

[10] Sinha RK & Roy AD & Kumar N & Mondal H & Sinha R. Applicability of ChatGPT in Assisting to Solve Higher Order Problems in Pathology. *Cureus*, 15(2), 1-9 (2023).

[11] Panda S & Kaur N. Exploring the Viability of ChatGPT as an Alternative to Traditional Chatbot Systems in Library and Information Centers. *Library Hi Tech News*, 40(3), 22-25 (2023).

[13] Kumar M & Kumar S & Kashyap PK & Aggarwal G & Rathore RS & Kaiwartya O & Lloret J. Green Communication in Internet of Things: A Hybrid Bio-inspired Intelligent Approach. *Sensors*, 22(10), 1-17 (2022).

[14] Kumar S & Rathore RS & Mahmud M & Kaiwartya O & Lloret J. Best-blockchain-enabled Secure and Trusted Public Emergency Services for Smart Cities Environment. *Sensors*, 22(15), 1-25 (2022).
[15] Ahn C. Exploring ChatGPT for Information of Cardiopulmonary Resuscitation. *Resuscitation*, 185, 1-2 (2023).
[17] Helberger N & Diakopoulos N. ChatGPT and the AI Act. *Internet Policy Review*, 12(1), 1-5 (2023).
[18] Dwivedi YK & Kshetri N & Hughes L & Slade EL & Jeyaraj A & Kar AK & Wright R. So What If ChatGPT Wrote It? Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy. *International Journal of Information Management*, 71(102642), 1-63 (2023).
[19] Lee IT. A Study on Characteristic and Application Method of Instrument to Measure Moral Identity. *Journal of Moral & Ethics Education*, 50, 1-27 (2016).
[20] Park GY. Moral Competence Test(MCT), Consciousness of Unification and Security, Moral, Psychology, C-index, Moral Development, Moral Reasoning. *Research Institute for National Security Affairs*, 60(3), 67-92 (2017).
[23] Lee SH. Is It Possible to be a Moral Artificial Intelligence? The Problem of Moral and Legal Responsibility in A. I. *Journal of Law and Politics Research*, 16(4), 283-302 (2016).
[24] Park GY & Jung B & Seo ES. Development of Board Game for Integrity Education using Moral Dilemma. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 9(12), 159-169 (2019).
[25] Narvaez D & Rest J. The Four Components of Acting Morally. *Moral Behavior and Moral Development: An Introduction*, 1(1), 385-400 (1995).
[27] Dehouche N. Plagiarism in the Age of Massive Generative Pre-trained Transformers(GPT-3). *Ethics in Science and Environmental Politics*, 2, 17-23 (2021).
[28] Hisan UK & Amri MM. ChatGPT and Medical Education: A Double-edged Sword. *Journal of Pedagogy and Education Science*, 2(1), 71-89 (2023).
[30] Cotton DRE & Cotton PA & Shipway JR. Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT. *Innovations in Education and Teaching International*, n2190148, 1-12 (2023).
[33] Taecharungroj V. What Can ChatGPT Do? Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data and Cognitive Computing*, 7(1), 1-10 (2023).
[36] Tlili A & Shehata B & Adarkwah MA & Bozkurt A & Hickey DT & Huang R & Agyemang B. What If the Devil is My Guardian Angel: ChatGPT as a Case Study of using Chatbots in Education. *Smart Learning Environments*, 10(1), 1-24 (2023).
[39] Zhou J & Ke P & Qiu X & Huang M & Zhang J. ChatGPT: Potential, Prospects, and Limitations. *Frontiers of Information Technology & Electronic Engineering*, n2300089, 1-6 (2023).
[41] Floridi L & Cowls J. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1), 535-545 (2019).
[43] Wong GK & Ma X & Dillenbourg P & Huan J. Broadening Artificial Intelligence Education in K-12: Where to Start?. *ACM Inroads*, 11(1), 20-29 (2020).
[46] Ashok M & Madan R & Joha A & Sivarajah U. Ethical Framework for Artificial Intelligence and Digital Technologies. *International Journal of Information Management*, 62, 102433-102440 (2022).
[47] Böhm S & Carrington M & Cornelius N & De Bruin B & Greenwood M & Hassan L & Shaw D. Ethics at the Centre of Global and Local Challenges: Thoughts on the Future of Business Ethics. *Journal of Business Ethics*, 180(3), 835-861 (2022).
[50] Farrow R. The Possibilities and Limits of XAI in Education: A Socio-technical Perspective. *Learning, Media and Technology*, 48(2), 266-279 (2023).

5.2. Books

[21] Kohlberg L. Development of Moral Character and Moral Ideology. University of Chicago (1964).
[22] Rest JR. Moral Development: Advances in Research and Theory. Praeger (1986).
[48] Pasquale F. New Laws of Robotics: Defending Human Expertise in the Age of AI. Harvard University (2020)

[49] Crawford K. The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University (2021).

5.3. Additional references

[3] Susnjak T. ChatGPT: The End of Online Exam Integrity?. ArXiv (2022).
[4] Lin J. ChatGPT and Moodle Walk into a Bar: A Demonstration of AI's Mind-blowing Impact on E-learning. SSRN (2023).
[5] Lee JA & Lee GJ & Kang Gy & Park YJ. From Speeches and Judgments to Composing… 'AI Chatbot' Swept the World in Two Months of Launch. Segeilbo (2023).
[12] Katyal N. Six Limitations of Conversational Artificial Intelligence. The Digital Edition of Dataquest Magazine (2020).
[16] Dayan P & Sahani M & Deback G. Unsupervised Learning. The MIT Encyclopedia of the Cognitive Sciences (1999).
[26] Zhuo TY & Huang Y & Chen C & Xing Z. Exploring AI Ethics of ChatGPT: A Diagnostic Analysis. ArXiv (2023).
[29] Akbar MA & Khan AA. Ethical Aspects of ChatGPT in Software Engineering Research. ArXiv (2023).
[31] https://www.nzherald.co.nz/ (2023).
[32] Mhlanga D. Open AI in Education, the Responsible and Ethical Use of ChatGPT towards Lifelong Learning. SSRN (2023).
[34] Vock I. ChatGPT Proves That AI Still Has a Racism Problem. The New Statesman (2022).
[35] Hacker P & Engel A & Mauer M. Regulating ChatGPT and Other Large Generative AI Models. ArXiv (2023).
[37] Weidinger L & Mellor J & Rauh M & Griffin C & Uesato J & Huang PS & Gabriel I. Ethical and Social Risks of Harm from Language Models. ArXiv (2021).
[38] https://openai.com/ (2022).
[40] https://www.theatlantic.com/ (2023).
[42] Krügel S & Ostermaier A & Uhl M. The Moral Authority of ChatGPT. ArXiv (2023).
[44] Albrecht J & Kitanidis E & Fetterman AJ. Despite Super-human Performance, Current LLMs are Unsuited for Decisions about Ethics and Safety. ArXiv (2022).
[45] https://time.com/ (2022).
[51] https://onezero.medium.com/ (2020).
[52] https://www.headmind.com/ (2023).

6. Appendix

6.1. Author's contribution

	Initial name	Contribution
Lead Author	JB	-Set of concepts ☑ -Design ☑ -Getting results ☑ -Analysis ☑ -Make a significant contribution to collection ☑ -Final approval of the paper ☑ -Corresponding ☑ -Play a decisive role in modification ☑
Corresponding Author*	GP	-Significant contributions to concepts, designs, practices, analysis and interpretation of data ☑ -Participants in Drafting and Revising Papers ☑ -Someone who can explain all aspects of the paper ☑

2023 Volume 8, 12-22 Submission: 2023/04/30, Peer review: 2023/05/30, Accepted: 2023/06/20, Published: 2023/08/30

Robotics & AI Ethics

Publisher: J-INSTITUTE
ISSN: 2435-3345

Website: j-institute.org
Editor: admin@j-institute.org

Corresponding author*
E-mail: hans.kim@pusan.ac.kr

DOI Address:
dx.doi.org/10.22471/ai.2023.8.12

Suggestions for Ethical Decision-Making Model through Collaboration between Human and AI

Hyunsoo Kim

Pusan National University, Pusan, Republic of Korea

Abstract

Purpose: *The purpose of this study is to explore and propose a model that allows humans and AI to collaborate in the process of making decisions about ethical issues. Due to AI's autonomy and mission performance capabilities, AI is sometimes viewed as an agent competing with humans. However, since the autonomy and mission performance capabilities of AI are applied at very diverse levels and areas, it is necessary to set certain categories and review their application. This study sought to reveal that more valid decisions can be made by collaborating between humans and AI in the category of ethical decision-making.*

Method: *This study uses methods of literature research and development research. First, using literature research to review various previous studies to understand the autonomy of AI in the relationship between humans and AI. Next, analyzing the meaning and characteristics of ethical judgment. Next, looking at a series of models that explain decision making. Second, using development research methods, for design and propose a model in which humans and AI appropriately collaborate in the process of making ethical decisions.*

Results: *The results of this study reveal the following points. First, the results of ethical decision-making by humans and AI involve greater responsibility and related issues than the results of general decision-making. Second, in order to solve these problems, it is necessary to utilize collective intelligence through collective decision-making and at the same time distribute responsibility. Third, as a public and collective entity functioning as a committee, humans become the subjects of final judgment and responsibility, and AI must play a role in actively and functionally assisting such judgment. Fourth, this decision-making process needs to be presented in the form of a model as a principle that can be applied to various specific cases.*

Conclusion: *The conclusion of this study suggests that effective and valid ethical decisions can be made through collaboration between humans and AI in the ethical communication process. And based on this, we present a collaboration model between humans and AI. This model consists of the following steps: First, AI should be actively involved in the process of exploring data sources, collecting data, storing data, and refining and analyzing data for ethical decisions. Second, ethical decisions based on this are made by a human community in the form of a committee as a group thinking process. Third, allow humans and AI to evaluate and exchange opinions on the results of these ethical judgments through mutual feedback and collaboration.*

Keywords: ***Artificial Intelligence, Ethical Decision-Making, Decision-Making Model, Human-AI Collaboration***

1. 1. AI Autonomy and the Relationship between Humans and AI

Discussions about autonomy function as a moral basis for human subjectivity and the dignity of humans as individuals. In that case, the inspection of AI's autonomy is not only the beginning of a discussion about the position of AI as a subject of action, but also the start of a discussion about the status of AI as a subject of action and judgment, and the resulting responsibilities and obligations of AI itself. . Therefore, examining AI with a focus on the concept of autonomy and examining the relationship between AI and humans based on this has the desired validity.

1.1. Mechanism of discussion on autonomy of AI

The most characteristic attribute of AI is its autonomy. Autonomy means the quality or characteristic of doing something according to one's own principles or controlling oneself and exercising restraint or self-control[1]. Therefore, autonomy means acting with free will as the subject of one's own actions[2]. Autonomy means a sense of ownership that allows one to live an independent life with high self-esteem by making choices and controlling a given task. This autonomy becomes an element that recognizes the subjectivity of AI at a certain level[3][4].

Autonomy is an important prerequisite for responsibility and rights. If so, the discussion on AI's autonomy can become a theoretical basis for norms to be applied in distributing responsibility and contributions related to AI's achievements[5]. Discussions regarding the autonomy of artificial intelligence can generally be divided into two parts: those that actively accept the autonomy of AI and those that approach it critically from a somewhat passive perspective.

The position that actively accepts the autonomy of AI generally focuses on the fact that AI makes independent decisions and acts. And we approach this in the form of analysis of the impact of AI's actions on humans. For example, the following studies acknowledge the autonomy of AI and, based on this, take the position of acknowledging AI as the subject of ethical judgment and action. Representative examples include research exploring the Ethical Implications of AI[6], Suggestions on structural Systems Design to Reflect Ethics in AI's Rules of Engagement Learning for Future Warfare[7], Forming Ethical AI as an Artificial Moral Agent in using Virtue Education Method[8], Building the AI Code of Ethics through Deep Learning and Big Data Based AI[9], etc.

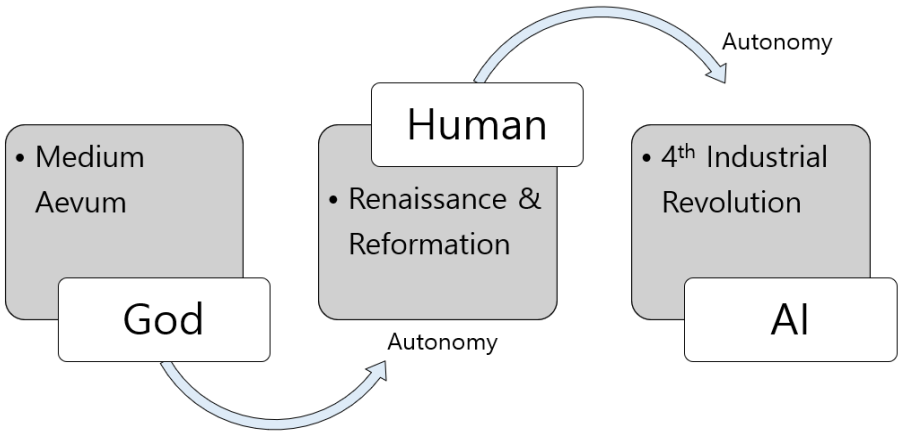
Meanwhile, those who view the autonomy of AI somewhat passively or critically tend to approach the autonomy of AI itself from the ontological aspect of philosophy. In order to provide morality to artificial intelligence, a study examined the harm and responsibility issues that may arise from artificial intelligence and pointed out that a normative approach is needed in AI ethics[10]. The autonomy of AI was evaluated as ethical impact agents, implicit ethical agents, A study that classified them into four stages or kinds of explicit ethical agents and full ethical agents, but argued that they cannot be viewed as moral machines or agents[11], explained that ethical control is necessary because Social Robots themselves have the risk of undermining human autonomy. Research[12], which analyzed the problems inherent in AI autonomy from three aspects: Driver-less Cars, Killer Robots, and Black-box decision making and legitimacy[13], are representative examples of such research.

The reason for discussing the autonomy of AI is to check whether it can have subjectivity and form an equal relationship with humans. This shows similar aspects to the beginning of modern times represented by the Renaissance. With the transition from the Middle Ages to the modern era, humans were understood as beings with reason who could make their own decisions and act. Accordingly, the idea that humans were created by God was maintained, while at the same time being liberated from dependence on God. This human free will is connected to basic human rights and to the ideas of liberalism and autonomy. If the idea of human autonomy is derived and developed from the God-centered order through this mechanism, the idea of AI autonomy can be similarly derived from the human-centered order. This is also embodied in discussions related to human free will in the field of scientific psychology[14].

Of course, the development of discussions on human autonomy and subjectivity from the God-centered order of the Middle Ages through the Renaissance cannot be completely equated with the current autonomy of the Fourth Industrial Revolution and AI. Nevertheless, explaining the subjectivity of humans and AI through the concept of autonomy is somewhat meaningful in

that it is a starting point for a commensurable ethical discussion necessary to discuss the equality and difference between humans and AI in the future. can be evaluated. This mechanism can be expressed as shown in <Figure 1>.

Figure 1. The mechanism of autonomy that supports the subjectivity and the dignity of an agent



1.2. Interrelationship between autonomous AI and humans

The most characteristic research on the autonomy of AI leads to examining what kind of relationship AI can have with humans. This can be looked at from two aspects. One is a discussion of the responsibility that comes with autonomy[15][16]. This is because autonomy does not presuppose self-indulgence or free riding. The other is a discussion about the scope and limits of autonomy[17][18]. This is because it is not permitted to infringe on or deny the autonomy of other subjects.

Autonomy is a key concept necessary to recognize the dignity of the subject. In particular, Kant explained the meaning of human dignity and respect for it. According to him, it is possible for humans to take responsibility for their actions because they decide their own actions. Therefore, if AI is understood as a rational, autonomous, and moral being, decisions made by artificial intelligence can be understood at the same level as human moral decisions[19].

Meanwhile, Kant emphasizes the importance of moral community. Of course, Kant's ethics primarily emphasizes the individual agent and his autonomy because it considers the source of morality to be inherent in the agent. However, the attribution of responsibility based on the autonomy of these actors operates as a prerequisite for moral evaluation and discussion of responsibility for community problems. Therefore, Kant's moral philosophy can be expanded from individualistic ethics to community-oriented[20].

The above discussion leads to the following explanation. First, if AI has a certain level of autonomy, it can make autonomous judgments and decisions accordingly. Second, if AI has autonomy, it can become a subject of moral judgment, and therefore has a certain level of equality with humans and can participate as a subject in the public forum of discussions on moral issues.

This discussion suggests that it is necessary to establish the mutual relationship between humans and AI as a relationship of coexistence. In general, research shows that the relationship between humans and AI should be understood in interaction rather than being evaluated by users and their output at a social level[21], and Studies that suggests that the benefits are greater than the costs in human-AI collaboration, and that human-AI synergy operates to a greater extent support this point[22].

The interaction between autonomous AI and humans can be examined in relation to Habermas' Public Sphere concept. This is understood as an area related to society, integration, and cultural transmission of members, as well as an area that manages society and produces materials[23]. However, content related to ethical decision-making can be seen as interpenetrating these two areas. Meanwhile, the Public Sphere can be seen as a space where active actors can freely communicate and make decisions through discourse[24]. At this time, if AI has the ability to participate in this public sphere and at the same time actually participates, there is room for recognition as functioning as a subject that interacts with humans at a certain level.

AI and humans with such autonomy will be able to contribute to achieving common goals by sharing certain roles, performing each other's tasks, and sharing the results. This common goal can exist in various forms, and it is requested to examine how Human-AI collaboration can be developed for ethical judgment and decision-making.

2. Understanding Ethical Judgment and Decision Making

2.1. Meaning and characteristics of ethical judgment

Ethical judgment or moral judgment refers to judging moral issues from a moral perspective among value judgments. However, although some of these ethical judgments require immediate judgment, there are also those that require evaluation over time for deliberation. Recent studies on AI's ethical judgment show a tendency to pursue AI's complete and autonomous judgment, which can be seen to be due to the following two factors. The first is to make ethical judgments completely automatically without human intervention. This is related to the goal of recognizing the independence of AI without human intervention. The second is to request immediate judgment in relation to the speed of ethical judgment. Therefore, it involves asking for quick judgment and action in a given situation without any discussion or deliberation.

The above two points ultimately guide us to set the direction of research as enabling AI to make independent ethical judgments. And this orientation works as a mechanism to design human morality into an algorithm and have AI learn it. This approach calls for analytical and empirical research on human morality. However, it is also a problem in the normative area, and it still contains problems in that the search for morality itself and the consideration of the ethical problems it brings are not complete even in human judgment.

Among various studies on human moral decisions, especially studies on brain science, it is revealed that human morality is an accumulation of various experiences and is also related to complexity and relationships[25]. Such ethical decisions are embodied in the form of ethical reasoning. Ethical reasoning is a form of reasoning that helps solve ethical problems and seek and secure the moral high ground. This ethical reasoning is based on ethical thinking, which has normative characteristics. In other words, reasoning based on ethical thinking and choices based on it are not based on what is right for the individual making the judgment. It must be a choice based on what is right for everyone who is directly and indirectly affected by the decision.

From this perspective, in order to resolve ethical conflicts, it is necessary to change or transition thinking from the level of intuitive judgment to the critical level of moral reasoning. When conducting such ethical thinking and reasoning, it is possible to carry out these ethical thinking and reasoning at a critical level based on an ethical perspective. At this time, the relationship between intuition and reasoning can be approached by understanding the difference between the human moral judgment mechanism and AI's moral reasoning mechanism.

Moral judgment based on intuition can be said to demonstrate the characteristics of human decision-making. The intuition that operates in moral judgment is not entirely dependent on rational thought or logical reasoning. This refers to the ability to generate direct knowledge or

understanding and reach decisions. To utilize this intuition, humans rely on practices, experiences, and judgments implemented in the past[26][27]. In comparison, moral judgment based on rational reasoning can be said to demonstrate the characteristics of an analytical approach to decision-making. This refers to the process of making judgments based on the amount and depth of information. In other words, decision-making relies on cognitive and information-driven processes as a direct result of intentional information collection and processing[28][29].

Intuition and reasoning examined in this way can be said to be two aspects that explain the operation of the human brain. AI seeks to track and learn from the human mind and thinking processes, thereby securing the ability to improve itself. However, these intuitions and reasoning are included in the complex aspects of human decision-making. Nevertheless, given that research related to the development of AI mainly relies on decision-making as a cognitive and information-driven process, even if AI forms intuition and reasoning together in the long term, it is an information processing technology that AI has strengths in the short term. There will be a need to focus on decision-making assistance technology based on it.

2.2. The need for collaboration between humans and AI in ethical decision-making

In order to utilize the advantages of intuition and reasoning in ethical decision-making, it is necessary for humans and AI to collaborate with each other rather than confront each other. This collaboration between humans and AI can be achieved through mutual recognition based on the social relationships formed between humans and AI[30]. First, if the social relationship between humans and AI is a universal egoistic relationship that preserves each other's interests in the relevant area and role, the required recognitional attitude is an attitude that recognizes rights, and the normative state achieved through this attitude is respect. should be set as the goal. Next, if special altruism is the goal, the recognitional attitude of love and the normative state of intimacy should be aimed. Lastly, if universal altruism is the goal, it should be aimed at a compassionate attitude called solidarity and a normative state of comradeship based on it. These divisions are as shown in <Table 1> below.

Table 1. Human-AI mutual recognition based on relationship.

Relationship / Recognition	Legal relationship / Moral relationship		
	Universal egoism	Special altruism	Universal altruism
Attitude of acceptance	Rights	Affection	Solidarity
Normative status	Respect	Belonging	Brotherhood

The collaborative intelligence between humans and AI constructed in this way basically aims to increase human efficiency and enable more valuable tasks to be performed more smoothly. However, such collaboration is often approached from the perspective of utility. In other words, when a decision must be made in the face of uncertainty, collaboration is useful in finding an answer to what decision to make. It is also intended to help you make the decisions necessary to answer the question of what information to use and how to use it.

However, in this case, the results of collaboration are usually approached from the perspective of decision theory. This goes through the following process. First, the evaluation of certain certainty is quantified by the decision probability of the decision maker. Next, we construct a utility theory to evaluate the value of the consequences resulting from a decision. And the actor who makes the decision deduces the expected utility based on the judgment probability, that is, the decision to maximize the expected utility, as a result of logic. Therefore, it is argued that making this decision is the rational and most appropriate decision under uncertainty for the

agent making the decision. Therefore, Preference, expressed as utility, is embodied in Decision Theory, which is formed through the combination of Probability Theory and Utility Theory [31].

There are points where ethical decision-making shows different characteristics from such general decision-making. This is because it must reflect the characteristics of ethical issues that must be considered in terms of qualitative factors that cannot be quantified or converted to quantity. In general decision theory, an agent is understood to be rational when it chooses an action that yields Maximum Expected Utility (MEU), averaged over all possible outcomes of a decision. However, it is pointed out that ethical decision-making, even if MEU is calculated, must be reviewed from multiple perspectives to determine whether there are inherently undesirable actions [32].

The most important thing in decision making will be the choice to calculate MEU. AI's inductive approach has strengths in these measurements and arithmetic calculations. It can be said to be an advantage of AI to extract various variables that need to be reflected, quantify and apply them, evaluate the final results of each choice, compare them, and present them as results. However, considering and judging various standards other than the MEU can be said to be a role that humans must perform as subjects of ethical judgment and an advantage that humans have. Therefore, human deductive judgment must be involved in determining what standards other than the MEU exist and in what situations they should be applied.

3. Example of Ethical Decision Making Model based on Human - AI Collaboration

3.1. Considerations in models related to ethical decision making

Ethical decision-making is based on making decisions based on an individual's moral judgment. However, in addition to the fact that such decisions can be influenced by subjective judgment, there is a risk that rational judgment may be difficult in practice due to various limiting factors. However, collective decision-making has higher accuracy and creativity in that the individuals who make moral judgments form a community to make judgments, and members of the community encourage and intellectually stimulate each other through interaction, and acceptability. It has the advantage of high satisfaction.

Such collective decision-making is appropriate in that it ensures the universality of ethical decision-making. Ethics is a standard of life in a personal sense, but it is also the order of human relationships. Therefore, the search for ethics as a standard that should be applied to all humans regardless of time and place has a positive meaning in that the more diverse the members participating in such discussions, the more active discussions can develop.

On the other hand, ethical issues are related to responsibility. It is not only impossible but also inappropriate to avoid the responsibility that comes with making a moral decision. However, on the other hand, when the burden of responsibility imposed by such a moral decision is large, the subject who makes the moral decision may avoid the burden of decision or responsibility because it is difficult to solve the problem of responsibility.

In ethical matters, collective decision-making serves the function of increasing the amount of deliberation related to this responsibility and reducing the psychological burden of responsibility. In this regard, it would be effective to appropriately utilize collective decision-making methods to respond to various unexpected ethical problems that will appear in future society.

A representative example of such collective decision-making was proposed in the medical field that deals with human life, focusing on the need for collective decision-making, especially in the process of making decisions related to the end of life [33]. The model of collective decision-making introduced in this medical field through the relationship and collaboration between

humans and AI is presented in <Table 2> below.

Table 2. Human-AI mutual recognition based on relationship.

Classification		Model		
Analysis	Subject	Human driven	Shared Decision making	AI autonomy
Exchange of information	Flow	One Way	Inter Action	One Way
	Direction	Human → Issue	Human + AI → Issue	AI → Issue
	Category or issues	Human interest	Various Issues	Extracted by AI
	Quantity	Minimum legal requirements	All factors involved in decision making	Most quantifiable and measurable factors
Deliberative thinking		Single / Community of human	Multi stakeholder	AI
Decision making body		Human	Human-AI coworking	AI

3.2. Structure of ethical decision making through collaboration between humans and AI

Research on collaboration between humans and AI is being developed in various forms, including research on the Data Communication Model[34], and research on the impact of AI on human cognition[35]. Based on these discussions, the structure of ethical decision-making through collaboration between humans and AI can be designed through certain procedures.

This procedure can be developed in a similar way to a general decision-making procedure, but has the following three characteristics. First, you will understand important ethical theories as a basis for decision-making, select an appropriate one, and apply it to the relevant issue. Second, throughout the decision-making process, humans and AI collaborate through functional linkage. Third, in the process of developing mechanisms related to decision-making, there are stages in which AI takes the lead in performing tasks and stages in which humans deliberate, evaluate, and decide, each with their own characteristics.

Meanwhile, the important ethical theories that should be considered as criteria for decision-making are generally presented in six categories: deontology, utilitarianism, common good, theory of Justice, virtue ethics, and care ethics, which are summarized in <Table 3> below.

Table 3. Examples of most important ethical theories

Subject	Core Idea
Deontology	The duty to respect others' rights and dignity
Utilitarianism	Emphasizing the consequences of our actions
Justice	Each person should be given their due which is interpreted as fair or equal treatment
Common Good	Life in community is a good in itself and people's actions should contribute to that life
Virtue	Actions consistent with ideal virtues that provides for the development of humanity
Care Ethics	Listen and respond to individuals in their specific circumstances

From an ethical perspective, judging and deciding on situations and issues that are subject to discussion through collaboration between humans and AI can be accomplished through the following procedures.

First, this is the step of identifying ethical issues. At this stage, the facts about the situation or problem that is the subject of discussion are checked and reviewed from an ethical perspective. There are many different perspectives on a situation or problem, and approaches can be made accordingly. For example, the same case can be approached from legal, political, economic, social, and cultural perspectives. However, this step sets the starting point and destination for viewing and interpreting these events from the perspective of ethics and sets the starting stage for conducting analysis.

Second, this is the stage of collecting information related to the facts related to the presented situation or issue. In order to make an ethical judgment based on an event, as much information as possible is needed. And after securing this information, it is necessary to set criteria for classifying the information and analyze and diagnose it accordingly. In order to approach a situation or issue from an ethical perspective, the information that must be considered to make an ethical judgment exists in the form of Big Data. Collecting such information and typing and diagnosing it to help with decision-making can be said to be specialized AI capabilities, and at this stage, AI will take the lead.

Third, it is a step to evaluate the analysis results and predict the results of alternative ethical actions. At this stage, given issues and analysis results are evaluated based on important ethical perspectives. At this time, in the process of learning about ethical perspectives, AI learns to make deductive judgments through Deep Learning, and sets certain standards for ethical choices and the resulting results according to each representative theory and quantitatively quantifies them. Activities to present results are carried out. At this stage, there is a need for AI to take the lead in carrying out the task in that AI takes the lead in learning, makes ethical judgments accordingly, analyzes the results, and presents them in parallel.

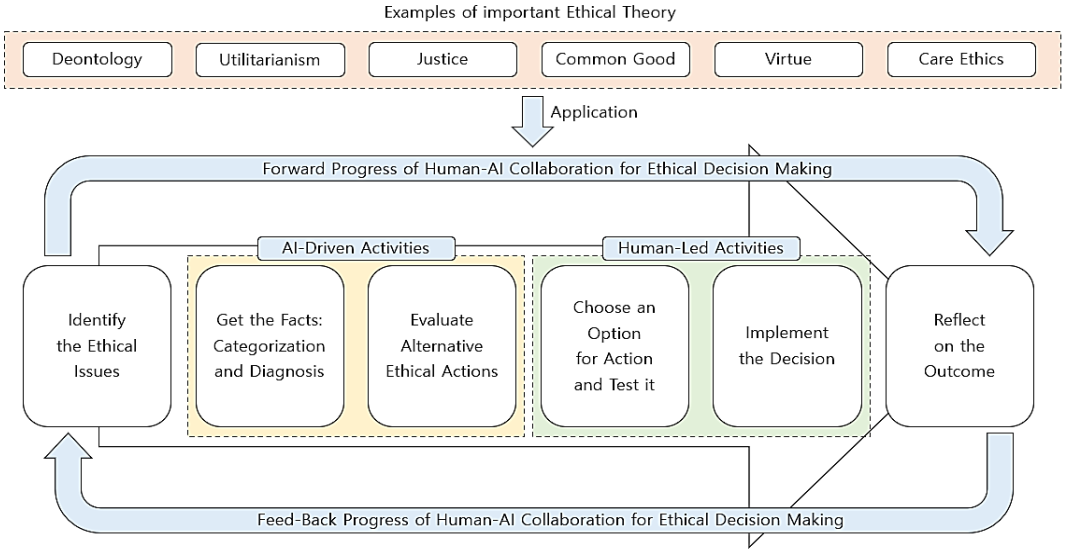
Fourth, it is the stage of choosing ethical behavior and testing it. At this stage, various matters to be considered when actually selecting the results analyzed and presented by AI earlier are evaluated and checked from a human perspective. This is the stage where corrections are made by human actors for AI's imperfections and information bias. At this stage, human experts evaluate AI learning and the analysis results presented by it, and if necessary, humans take the lead in correcting unexpected problems through override and correcting Deep Learning. This is the stage where it becomes possible.

Fifth, from an ethical perspective, it is the stage of making a choice or decision through careful deliberation and applying it in practice. Making an ethical judgment is also an existential choice, but in order to apply it more effectively, it is necessary to make a judgment within a collective decision-making system in which humans and AI collaborate, as discussed above. What is important at this time is that it is inevitable that human actors will ultimately be the ones responsible for the various stakeholders affected by such judgments. If so, the result is that the role of humans as the subject of these responsibilities and obligations plays an important role in ethical decisions and their accompanying results.

Sixth, this is the final stage of reflection and evaluation of this overall procedure and its results. Some ethical decisions arise from unique and special circumstances. But at the same time, it serves the function of foreshadowing other similar ethical issues. If so, there is a need to conduct an active and complex analysis of various factors, including the process of these discussions, the roles of the decision-making entities involved in the process, and reflection on the results. This analysis will provide data that will enable more sophisticated and refined decisions to be made on issues that require further ethical judgments that will arise in the future.

Each of the above procedures is presented in the form of a model as shown in <Figure 2>.

Figure 2. Ethical decision-making model through human-AI collaboration



4. Conclusion

Collaboration between humans and AI is expected to unfold in various forms. However, such collaboration appears to be actively accepted, especially in terms of science, technology and economics, but in areas related to ethical discussions, there is still a passive understanding of AI's judgment and responsibility. Even if AI is not recognized as a member of the future society on an equal level with humans, it can be accepted that AI is an entity that lives together with humans in the future society. If so, AI may be able to contribute to ethical judgment at a certain level even if it is not the subject of perfect judgment.

AI's collaboration with humans in ethical decision-making has the advantage in the following three aspects: First, case studies utilizing AI's Deep-Learning through collaboration between AI and humans on representative important ethical theories and Learning will be carried out. This will lead to learning of basic data related to ethical decision-making that will be developed in the future. Second, AI utilizes the function of deliberation rather than immediate decision-making. This can broaden the scope of thinking about AI's function as an ethical agent. Third, it complements the intuitive judgments made by human decision makers and allows for more in-depth inspection of the results of ethical judgments.

Ethical judgment is the basis of ethical behavior. Therefore, research on ethical judgment through collaboration between AI and humans will lead to future research on AI's ethical sensitivity, ethical motivation, and ethical behavior.

5. References

5.1. Journal articles

[1] O'Neill T & McNeese N & Barron A & Schelble B. Human-autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors*, 64(5), 904-938 (2022).
[2] Ryan RM & Deci EL. Self-regulation and the Problem of Human Autonomy: Does Psychology Need Choice, Self-determination, and Will? *Journal of Personality*, 74(6), 1557-1586 (2006).
[3] Formosa P. Robot Autonomy vs. Human Autonomy: Social Robots, Artificial Intelligence (AI), and the Nature of Autonomy. *Minds and Machines*, 31(4), 595-616 (2021).

[4] Lawless WF & Mittu R & Sofge D & Hiatt L. Artificial Intelligence, Autonomy, and Human-machine Teams - Interdependence, Context, and Explainable AI. *AI Magazine*, 40(3), 5-13 (2019).
[5] Ilachinski A. Artificial Intelligence and Autonomy: Opportunities and Challenges. *Center for Naval Analysis*, 10-2017 (2017).
[6] Park G & Bae M. A Case Study on Current Issues in Artificial Intelligence and its' Ethical Implications. *Robotics & AI Ethics*, 7(2), 47-56 (2022). [Read More]
[7] Kim HS. Instructional Systems Design to Reflect Ethics in AI's Rules of Engagement Learning for Future Warfare. *Robotics & AI Ethics*, 6(4), 64-74 (2021). [Read More]
[8] Kim H. Approaches to Forming Ethical AI as an Artificial Moral Agent: Suggesting Virtue Education Method through Comparison of Top-down and Bottom-up Approaches. *Robotics & AI Ethics*, 6(2), 44-51 (2021). [Read More]
[9] Kim HS. Suggestion of Building the AI Code of Ethics through Deep Learning and Big Data based AI. *Robotics & AI Ethics*, 6(1), 29-34 (2021). [Read More]
[10] Hagendorff H. The Ethics of AI Ethics: An Evaluation Guidelines. *Minds and Machines*, 30, 99-120 (2020).
[11] Moor J. Four Kinds of Ethical Robots. *Philosophy Now*, 72, 12-14 (2009).
[12] Mezrich JL. Is Artificial Intelligence (AI) a Pipe Dream? Why Legal Issues Present Significant Hurdles to AI Autonomy. *American Journal of Roentgenology*, 219(1), 152-156 (2022).
[13] Chesterman S. Artificial Intelligence and the Problem of Autonomy. *Notre Dame Journal on Emerging Technologies*, 1(2) , 210-250 (2020).
[14] Guyer P. Kant on the Theory and Practice of Autonomy. *Social Philosophy and Policy*, 20(2), 70-98 (2003).
[15] Tigard DW. Responsible AI and Moral Responsibility: A Common Appreciation. *AI and Ethics*, 1(2), 113-117 (2021).
[16] Dastani M & Yazdanpanah V. Responsibility of AI Systems. *AI & Society*, 38(2), 843-852 (2023).
[17] Coeckelbergh M. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explain ability. *Science and Engineering Ethics*, 26(4), 2051-2068 (2020).
[18] Constantinescu M & Voinea C & Uszkai R & Vică C. Understanding Responsibility in Responsible AI. Dianoetic Virtues and the Hard Problem of Context. *Ethics and Information Technology*, 23, 803-814 (2021).
[19] De Cremer D & Kasparov G. The Ethical AI - Paradox: Why Better Technology Needs More and Not Less Human Responsibility. *AI and Ethics*, 2(1), 1-4 (2022).
[20] Stroud SR. Kant on Community: A Reply to Gehrke. *Philosophy & Rhetoric*, 39(2), 157-165 (2006).
[21] Salvini P & Laschi C & Dario P. Design for Acceptability: Improving Robots' Coexistence in Human Society. *International Journal of Social Robotics*, 2, 451-460 (2010).
[22] Sundar S. Rise of Machine Agency: A Framework for Studying the Psychology of Human-AI Interaction (HAII). *Journal of Computer-mediated Communication*, 25(1), 74-88 (2020).
[23] Garnham N. Habermas and the Public Sphere. *Global Media and Communication*, 3(2), 201-214 (2007).
[24] Benson R. Shaping the Public Sphere: Habermas and Beyond. *The American Sociologist*, 40(3), 175-197 (2009).
[25] Funk CM & Gazzaniga MS. The Functional Brain Architecture of Human Morality. *Current Opinion in Neurobiology*, 19(6), 678-681 (2009).
[26] Haidt J & Graham J. When Morality Opposes Justice: Conservatives Have Moral Intuitions That Liberals May Not Recognize. *Social Justice Research*, 20(1), 98-116 (2007).
[27] Graham J & Haidt J & Rimm-Kaufman SE. Ideology and Intuition in Moral Education. *International Journal of Developmental Science*, 2(3), 269-286 (2008).
[28] Bucciarelli M & Khemlani S & Johnson-Laird PN The Psychology of Moral Reasoning. *Judgment and Decision Making*, 3(2), 121-139 (2008).
[29] Ditto PH & Pizarro DA & Tannenbaum D. Motivated Moral Reasoning. *Psychology of Learning and Motivation*, 50, 307-338 (2009).

[30] Hohfeld WN. Fundamental Legal Conceptions as Applied in Judicial Reasoning. *The Yale Law Journal*, 26(8), 710-770 (1917).
[31] Klibanoff P & Marinacci M & Mukerji S. A Smooth Model of Decision Making under Ambiguity. *Econometrica*, 73(6), 1849-1892 (2005).
[32] Lindley DV. The Role of Utility in Decision-making. *Journal of the Royal College of Physicians of London*, 9(3), 225-230 (1975).
[33] Elwyn G & Edwards A & Kinnersley P. Shared Decision Making and the Concept of Equipoise: The Competences of Involving Patients in Healthcare Choice. *British Journal of General Practice*, 50(460), 892-899 (2000).
[34] Kim C. A Study on the Development of Data Communication Module within Construction. *Robotics & AI Ethics*, 6(4), 13-22 (2021). [Read More]
[35] Hu Q & Lu Y & Pan Z & Gong Y & Yang Z. Can AI Artifacts Influence Human Cognition? The Effects of Artificial Autonomy in Intelligent Personal Assistants. *International Journal of Information Management*, 56, n102250 (2021).

6. Appendix

6.1. Author's contribution

	Initial name	Contribution
Author	HK	-Set of concepts ☑ -Design ☑ -Getting results ☑ -Analysis ☑ -Make a significant contribution to collection ☑ -Final approval of the paper ☑ -Corresponding ☑ -Play a decisive role in modification ☑ -Significant contributions to concepts, designs, practices, analysis and interpretation of data ☑ -Participants in Drafting and Revising Papers ☑ -Someone who can explain all aspects of the paper ☑

2023 Volume 8, 23-36 Submission: 2023/04/30, Peer review: 2023/05/30, Accepted: 2023/06/20, Published: 2023/08/30

Robotics & AI Ethics

Publisher: J-INSTITUTE
ISSN: 2435-3345

Website: j-institute.org
Editor: admin@j-institute.org

Corresponding author*
E-mail: skcho@ikw.ac.kr

DOI Address:
dx.doi.org/10.22471/ai.2023.8.23

Soft Power in Northeast Asia, Using AI in Information Warfare

Sunggu Jo

Kyungwoon University, Gumi, Republic of Korea

Abstract

Purpose: *Conflicts between Korea, the United States, Japan, North Korea, China, and Russia continue along with the competition for supremacy between the United States and China, and conflicts between South and North Korea and between China and Taiwan continue as the Korean Peninsula is an area with a high possibility of military conflict. However, even in North Korea, China, and Russia, which are closed countries, the influence of the soft power of cultural content such as dramas through the Internet is bringing about changes in collective sentiment even in closed countries. Due to these phenomena, Northeast Asia, which has been confronted with military power, is facing a new phase, and we wanted to discuss the use of AI technology in information warfare by the soft power of the Intelligence Agency.*

Method: *For the expansion of AI use and research on information warfare, the historical cases of Northeast Asia were analyzed, and the evolution of literature and media was reviewed to understand the phenomenon of artificial intelligence (AI) after the 4th industrial revolution, and the themes were selected. In addition, related data were collected and reviewed, and an attempt was made to theoretically establish the research results academically.*

Results: *1. Northeast Asia is undergoing a transition from order based on hard power through military power to soft power based on cultural content. Just as the spread of culture has expanded faster and deeper the more it is controlled by the state, many researchers in Northeast Asia sympathize with the collapse of the system when asked how long such surveillance and control by dictatorships such as China and North Korea will be possible. According to this phenomenon, the influence of the power of culture on society was analyzed.*

2. Artificial intelligence (AI) learning information will adversely affect sound soft power due to manipulated information and biased algorithm learning data. Due to this loophole, the Intelligence Agency will launch an information war using artificial intelligence (AI) technology that suits its own interests. In addition, the Internet will accelerate the propagation speed of distorted soft power and penetrate deeply into human life. Therefore, the Intelligence Agency is expected to analyze the influence of this distorted soft power on its country and start blocking and defending against attacks.

Conclusion: *1. The legal system before the advent of AI is expected to be modified or supplemented by more than 50% after the advent of AI. In accordance with this paradigm shift, the authority of the Intelligence Agency in the information warfare of AI was divided into the right to investigate, the right to investigate, and the right to operate. 2. In response to the threat of using artificial intelligence (AI) in information warfare, the government of the country not only expanded the size of the Intelligence Agency but also proposed a hybrid structure of cooperation with the private sector, that is, a model of 'hybrid defense'. Lastly, in 1983, when tensions between the US and the Soviet Union were in the Cold War, the state-of-the-art scientific equipment, a satellite for detecting nuclear missiles, recognized the US ICBM launch warning. In response, Lieutenant Colonel Stanislav Petrov, commander of the Watch Command of the Soviet Air Defense Force, determined through human intuition that this was a computer error. It reexamined the case of preventing World War III by judging computer errors through human intuition, not judgment of scientific equipment, and suggested ethical issues in information warfare.*

Keywords: ***AI, Information Warfare, Intelligence Agency, Soft Power, Ethical Suggestions***

1. Research Purpose

As for the threat to humans, they developed rice farming during the Neolithic Age, started collective life, and formed clans and tribal states, and the threat of neighboring tribal states formed the concept of security[1]. In particular, the traditional security background of the Northeast Asian region has a historically complex relationship. It can be understood from the past wars and invasions, and many conflicts still exist as a bumper zone in the Asia-Pacific region[2][3].

In the past, the United Kingdom and the Russian Empire, the United States and the Soviet Union, and now the United States and China's hegemony competition is fiercely occurring in Northeast Asia[4][5][6][7]. In the course of this struggle for supremacy, many conflicts and wars have occurred in Northeast Asia, but the new Cold War structure between Korea, the United States and Japan vs. North Korea, China and Russia is still maintained[8][9].

Due to this, the Korean Peninsula serves as a buffer zone in the Asia-Pacific region, and there is a high possibility of military conflict in this region[10], the international community is making various efforts to mitigate these conflicts and promote regional stability. However, the prevailing opinion is that Northeast Asian issues cannot be resolved through dialogue and negotiations any longer[11][12].

However, recently, the absorption of cultural contents of the Korean Wave in North Korea, China, and Russia, which are closed countries, is expanding very rapidly due to the spread of the Internet. Exchanges at the private level, not the government level, from hard power through military power in the past to soft power through cultural content are rapidly increasing[13][14].

In addition, North Korea's possession of nuclear weapons following Russia and China in Northeast Asia has created a justification for 'balance' in the military alliance between South Korea, the United States, and Japan[15], and this 'balance of fear' is suppressing escalation of war.

However, at least about 3 million people died of starvation due to a famine that occurred from 1994 to 2000 in North Korea, and as the government's rationing system was suspended, the existing "communism to eat well and live well" politics disappeared. Residents who escaped North Korea are voicing their voices that it has now become a pseudo-religious dictatorship that demands "unconditional loyalty to Kim Il-sung, Kim Jong-il and Kim Jong-un"[16].

Therefore, fearing the spread of dissatisfaction with the North Korean regime among the people, the North Korean regime strengthened surveillance and control of the people. However, Korean dramas that spread rapidly in the 'jangmadang' are watched among North Koreans amidst fear and surveillance that they could be shot when discovered, proving the effect of soft power in changing the collective emotions of North Koreans more deeply[17].

Northeast Asia is undergoing a transition from order based on hard power through military power to soft power based on cultural content. This study dealt with the role of an Intelligence Agency using artificial intelligence (AI) in the dissemination of culture corresponding to soft power. In addition, with the advent of AI, the role of the Intelligence Agency and should evolve into a judicial basis, and through this discussion, a model for changes in the right to 'Investigative Authority', 'Prosecutorial Authority', and 'Operational Authority' rights of the Intelligence Agency was presented. And as a way to lead the technological development of artificial intelligence (AI) use in terms of efficiency in information warfare, a hybrid structure of cooperation with the private sector, that is, a model of 'hybrid defense' was presented.

As such, Northeast Asia has faced a new phase due to soft power, and it is expected that the use of AI technology in information warfare will gradually expand due to the influence of this soft power. In this study, an in-depth discussion was conducted on the use of AI technology in information warfare.

2. Transition from Hard Power to Soft Power

Korean dramas are popular in China and North Korea. In particular, the dramas 'Goblin' and 'Descendants of the Sun' are the most popular dramas in China and North Korea[18][19]. In particular, these dramas are illegally circulated in North Korea to avoid government surveillance, and North Koreans watch them on DVD or USB. Because these dramas are so popular, some North Koreans are looking for ways to watch these dramas by defecting to South Korea or through exchanges with South Koreans[20].

In the past, the effect of control through social surveillance was possible to some extent, but the more culture is controlled, the faster and deeper it expands. As such, the question arises as to how long North Korea's surveillance and control will be possible. Many researchers in Northeast Asia place weight on regime collapse and sympathize with regime change.

In August 2023, it is the situation of bomb terrorism in Pyongyang reported by multiple media outlets in Korea as a basis for signs that may lead to such internal unrest. North Korea is different from the general country we think of, and representatively, North Korean residents do not have freedom of passage, so people who do not live in Pyongyang, the capital of North Korea, cannot enter Pyongyang. These measures are taken from the viewpoint of protecting Pyongyang from the point of view of security for Kim Jong-un, and in reality, the lives of residents in Pyongyang and non-Pyongyang regions show too great a difference It is intended to prevent.

However, about the situation in which bomb terrorism occurred in Pyongyang and casualties occurred, and that a TF team was newly established to search for disgruntled people An official from the National Intelligence Service (NIS), who attended the South Korean National Assembly, explained, “There are unstoppable complaints and collective protests against the policies of the Kim Jong-un family and the party, centered on each generation in North Korea.” This means that collective sentiment in Pyongyang, North Korea is also changing.

Figure 1. Popular Korean dramas in North Korea: 'goblin' and 'descendants of the sun'.

Soft power is the ability to move and influence other countries through culture, which is much more effective than traditional military influence and has the characteristics of maintaining influence over a longer period of time[21][22].

In Northeast Asia, Korea's K-pop, K-drama, and K-beauty are examples of global soft power influence. Cultural contents were produced in Korea in Northeast Asia and shared and disseminated by people around the world through YouTube and Netflix. Recently, it has expanded to include not only information transmission using the Internet and social media, but also cultural experiences using virtual reality technology and language exchange using artificial intelligence technology. In particular, the expansion of online life due to the Corona 19 virus over the past few years seems to have contributed greatly to the interest in the Korean Wave.

In addition, the fan community plays a bigger role in expanding soft power, and fans freely communicate and share content online, thereby spreading their influence. For example, "ARMY", a BTS fan club, actively uses SNS to make comments about social issues, and through this, the social influence is considerable.

It is clear that strengthening soft power plays a very important role in the strategic development and security of the country, but it also has a dilemma that makes it less effective if the state directly intervenes in the growth of soft power.

Korean culture has grown a variety of pop culture contents such as music, dramas, movies, and games, and these contents have gained worldwide popularity and spread to various countries and cultures. Especially in Northeast Asia, experiences through music, dramas, and movies promoted the dissemination of the Korean language and led to tourism in Korea. South Korea's Incheon International Airport is positioning itself as a hub airport in Northeast Asia. In recent years, not only cutting-edge technology from global companies such as Samsung Electronics, but also Hallyu culture such as BTS has received worldwide attention, and cultural influence has grown, making Seoul the choke point for information activities in various countries. These purposes include not only technology leakage, but also the purpose of collecting trends in Korea and driving public opinion in a direction that is in line with the country's interests [23].

Table 1. The influence of cultural forces on society.

Division	Detail
Identity formation	Culture is the elements of culture, such as language, tradition, mythology, art, and music, that shape and reinforce the identity and self of individuals and groups.
Value formation	Culture shapes social norms and moral values, and influences individual and group behaviors and attitudes.
Communication expansion	Culture increases the efficiency of communication through a common language, signs, and symbol systems, and helps to understand and accept other cultures.
Social change	Acceptance and dissemination of new ideas, values, and technologies occur through culture, and cultural changes lead to social, political, and economic changes.
International influence	The attractiveness and influence of culture helps to form relationships with other countries and expand its international influence
Economic creation	Cultural industries, such as tourism, art works, music, films, and festivals, promote economic activities and create jobs, as well as develop local economies and tourism industries.

3. AI Threats and Intelligence

3.1. Risks of AI learning information

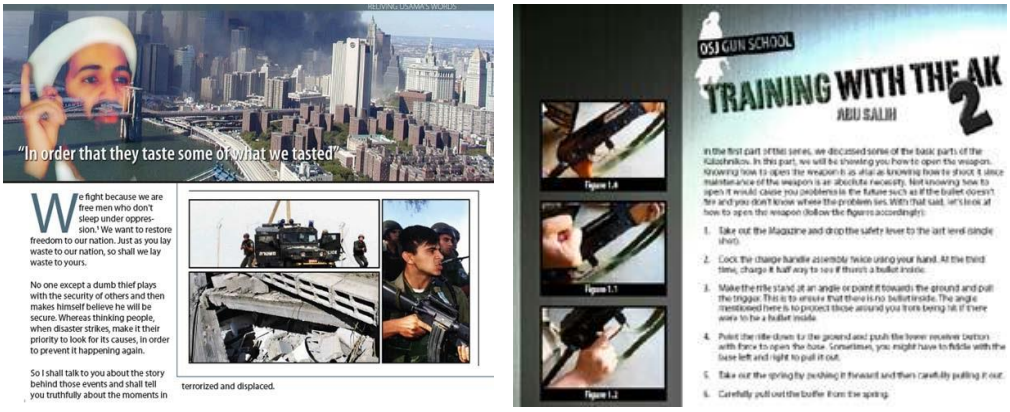
When artificial intelligence operates autonomously, inaccurate decisions are made by AI with different ethical standards from humans, but responsibility and regulation are difficult in information warfare. The continuous use of artificial intelligence (AI) technology in information warfare that has a psychological impact can completely destroy human sociality, leaving the interests of the country. You can understand the situation in more detail if you recall the 'RoboCop', which was introduced as a movie in the 1990s.

For example, with the gradual expansion of artificial intelligence (AI), artificial intelligence (AI) replaces traffic police enforcement work, and artificial intelligence (AI), that is, when machines regulate and control humans becomes legally common in more fields. At that moment, human dignity is also expected to be judged by artificial intelligence (AI).

Meanwhile, with the Fourth Industrial Revolution, the use of artificial intelligence (AI) technology in information warfare will gradually increase. However, artificial intelligence (AI) technology based on algorithm learning data biased by manipulation of information and fake news also affects healthy soft power. Against this background, the intelligence departments of each country will use artificial intelligence (AI) technology that meets their own interests to open information warfare [24].

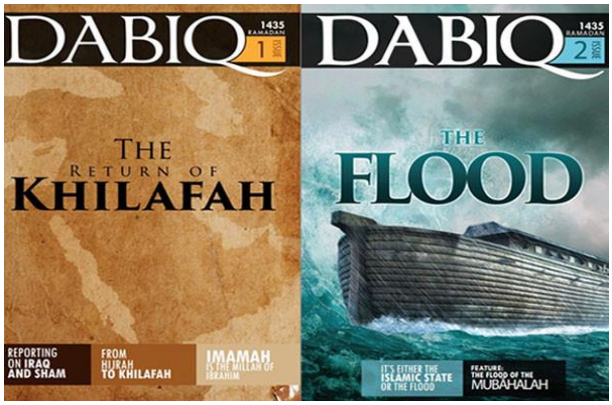
In addition, artificial intelligence (AI) learns harmful information in some cases, but INSPIRE, published in American English by Al Qaeda hardliners, publishes and distributes bomb-making methods online in PDF format on the Internet, so there is no realistic way to prevent. And anyone in the world can see INSPIRE if they want[25]. INSPIRE inspired countless potential terrorists around the world who felt relative poverty, and cases linked to actual terrorism were revealed through the Intelligence Agency. Al-Qaeda as well as ISIS are publishing DABIO similar to INSPIRE. When such dangerous publications are learned from generative artificial intelligence (AI) data such as ChatGPT, expanded and reproduced, and provided information to an unspecified majority, society will become more confused and dangerous.

Figure 2. INSPIRE published by Al-Qaeda[26].



Note: The most notable article on INSPIRE is a serial titled Open source Jihad. The magazine introduced it as a corner that allows readers to learn how to terrorize at home and jump into the holy war. Detailed methods of terrorism, such as how to easily make a bomb in the kitchen and how to blow up a building, are explained with easy-to-understand pictures. In our Fall 2010 issue, we introduced how to "sweep" people out with a pickup truck (a small truck without a lid on the cargo box). He asked for blades to be attached to the front and rear of the vehicle, along with a "friendly" explanation that a butcher's knife or a thick iron plate would be suitable. Regarding the location of the terror attack, he said that he likes a crowded but narrow space, and that a pedestrian-only space in the city center is the most ideal. It is explained that it is suitable as a "martyrdom operation" because it will be difficult to escape safely once it is carried out.

Figure 3. DABIO published by ISIS[27].



Note: Following INSPIRE, even ISIS has published an English promotional webzine, focusing on fostering 'autogenous terrorists'. ISIS evaluated the attack on Monis, who took hostages in Sydney, Australia in Dabiq, as "a war against the crusaders (western countries)." "Instead of coming to Khalifa's territory, he waged war alone on the streets that Western powers considered safe, and caused great terror in Australia simply by holding a hostage in a cafe with a single gun. he will receive the grace of god."

3.2. Globalization of information and expansion of artificial intelligence (AI)

The Internet has strongly promoted the 'internationalization' of Northeast Asian culture, and the speed of propagation of threats has also accelerated. Now, with cyberspace as a mediator and human life as a dependent variable, the comprehensive influence of artificial intelligence (AI) cannot be ignored.

Activities for the information collection or investigation in information warfare rapidly expanded to activities for attacking and defending by analyzing and examining information obtained through exchanges between cultures and evaluating the impact of cultural exchanges on the interests of one's own country.

In particular, the emergence of the Internet not only accelerates information, but also narrows the distance between countries and drives integration, and the use of the Internet is being utilized in various aspects. The Internet is spreading faster due to improvements in network infrastructure, high-speed Internet connectivity, cache, and content delivery networks (CDNs), cloud computing, content compression, and optimization, and advances in mobile technology.

Here, artificial intelligence technology has begun to replace the role that humans have played and is expected to penetrate more deeply into human life[28][29][30].

3.3. Role of the intelligence agency

Controversies such as the US National Security Agency (NSA) surveillance of civilians are a dilemma in which the intelligence departments of the free camp, excluding North Korea, China, and Russia, are experiencing a conflict of values between 'personal information' and 'threat information'[31]. However, as the number of US intelligence agencies officially reaches 16, the countries of the free camp are also doing their best in information warfare.

In liberal countries, this phenomenon is argued by some politicians and media as if the Intelligence Agency intervened in domestic politics.

However, such a situation will always be premised on the complete absence of aggressive political intervention in the hostile state.

In the current democratic political system, when the influence of collective psychology by non-combat factors such as cultural content increases its dominance throughout society, it is necessary to seriously recognize the loophole in the circulation structure in which social public opinion is connected to politics through elections.

This represents the beginning of ‘comprehensive security’, which has gone from physical threats in the past to cyber threats. Comprehensive security is a concept that protects the people based on information power in response to various unpredictable attacks and uncertainties by enemy countries.

Table 2. The use of artificial intelligence (AI) technology in information warfare.

Division	Detail
Information collection and analysis	Artificial intelligence efficiently derives information by collecting and analyzing a large amount of data from various sources, and can identify enemy activities or risk factors by analyzing satellite images, communication information, and open source information.
Language analysis and decoding	Artificial intelligence can be used to analyze and decipher multilingual documents, decipher encrypted messages or find important information or patterns in text data.
video and audio analysis	AI can extract useful information by analyzing video and audio data collected through drones, cameras, and microphones, and can detect enemy locations or analyze suspicious behavior.

Automated analysis and reporting	Artificial intelligence can quickly process complex data through automated analysis functions and create reports by summarizing analysis results.
predictive modeling	Artificial intelligence can use existing data and machine learning to predict likely events in a specific region or situation.
Autonomous exploration or search technology	Artificial intelligence can be used to perform dangerous areas or missions using unmanned drones or robots, and can be used to locate enemies or explore dangerous areas.
Signal analysis	AI can analyze radio communications or radar signals to detect enemy behavior or equipment.

Information warfare is strategically utilized by a country or organization to achieve goals by collecting, analyzing, and using information. Recently, artificial intelligence (AI) technology has had a great impact on information warfare. By analyzing the social media conversations and reactions of intelligence agents, you can identify and predict the emotions and sensitivities of your group to establish strategies. At this time, through AI deepfake detection technology, video, image, text, voice, etc. can be analyzed to identify false information, and another threat can be predicted by analyzing the pattern of identified information.

In addition, in the not-too-distant future, the VUCA environment is expected to accelerate further as methods combined with advanced science and technology, such as highly developed quantum computing, biotechnology and genome editing, robotics and drones, and artificial neural network reinforcement.

5. Discussion

5.1. Judicial powers in AI information warfare

5.1.1. Investigative authority

In information warfare, the investigation authority is divided into human information (HUMINT) and SIGINT (SIGINT), which captures signals using electronic equipment such as satellite photography or wiretapping. Sijint is classified into electronic information (ELINT), technical information (TECHINT), and communication content (COMINT), and means the start of information warfare.

Although intelligence gathering is being widely expanded around the world, it has become politically controversial in free countries, except for non-human rights countries such as North Korea and China, but with the advent of AI, such political debates are expected to become increasingly irrelevant.

Therefore, it is necessary to establish an active diversification strategy for the investigation right of the Intelligence Agency, and if the intelligence departments of each country indiscriminately expand artificial intelligence (AI) in information warfare, as no region can be exempt from changes in the global environment, not only their national interest but also human We must be stern in recognizing that the adverse effects circulate and affect our own country again.

5.1.2. Prosecutorial authority

In information warfare, the prosecutorial authority is used for the purpose of investigating espionage and terrorist activities in domestic counterintelligence activities, prosecuting through the prosecution, submitting evidence to the court, and blocking and defending against enemy attacks through judicial processing.

Currently, in liberalized countries, it is meaningless to divide the Intelligence Agency's duties into overseas and domestic, or online and offline, due to travel liberalization and the development of the Internet. Even soldiers and civilians are not controlled as in wartime, so the meaning of special distinction is diminished.

Therefore, information warfare strategies should be dealt with in a comprehensive concept by dividing them into combative and non-combatant factors from indiscriminate attacks using the enemy's artificial intelligence (AI). In addition, it is necessary to seek changes to an integrated strategy that improves the timeliness and security of investigations through diversification strategies.

The integrated strategy simplifies the decision-making process to enable timely and thorough attack blocking and defence on threat information, and also enables early identification of threats through prosecutorial authority. In addition, security can be secured in preserving evidence collection, which can prevent additional threats. Synergy can be maximized by strengthening the cooperative system of related organizations through the investigation command of the Intelligence Agency through comprehensive investigation data.

5.1.3. Operational authority

Operational authority is the authority used by the military or intelligence department, and means the authority to carry out necessary operations or take urgent measures to maintain national security, and all acts performed for the purpose of national security fall within the scope of operational authority.

Operational authority represents an important role for the government as it includes the authority to make decisions related to national security. However, under the Diplomatic Immunity and the New York Convention on Diplomatic Relations, military officers and intelligence agents disguised as diplomats dispatched overseas are not subject to judicial punishment through legal immunity in the sending country. This creates a dilemma between national interests and diplomatic relations. In other words, electronic information warfare, including the use of artificial intelligence (AI), does not follow the procedures of investigation, prosecution, and trial based on domestic law. It is also distinguished from the use of compulsory power following the issuance of a warrant in a domestic court.

The aspect of today's information warfare has changed with the use of advanced science and technology and information technology. In the future, as satellite information and GPS technology using artificial intelligence (AI) become the core of information warfare, the dependence on the use of GPS or GNSS is increasing not only in military facilities but also in the private sector, transforming into an aspect of small-scale operations in the city.

This is to neutralize the enemy's weapon function through low magnetic spectrum control. Operation using Jamming, Spoofing, Mikoning, etc. in the civilian area of the city is not only the enemy's weapon, but also its own civilian facility. It includes a judicial basis for establishing an operational area.

In this regard, if we look at the case of MI6 introduced by the British Guardian, Article 7 of the MI6 (Secret Intelligence Service Act) has the legal basis that ‘there shall be no legal liability if an act taking place overseas is carried out with the permission of the Minister.’ It is said that requests to apply Article 7 to covert operations carried out by the UK around the world have increased rapidly since the September 11 at-tacks, with an average of 500 ministerial signatures per year.

Breaking away from the practice of secretly and implicitly in information warfare in the past, establishing judicial grounds and standards for 'Operational Right' even in non-disclosure will prevent attacks from comprehensive threats resulting from the expansion of the use of artificial

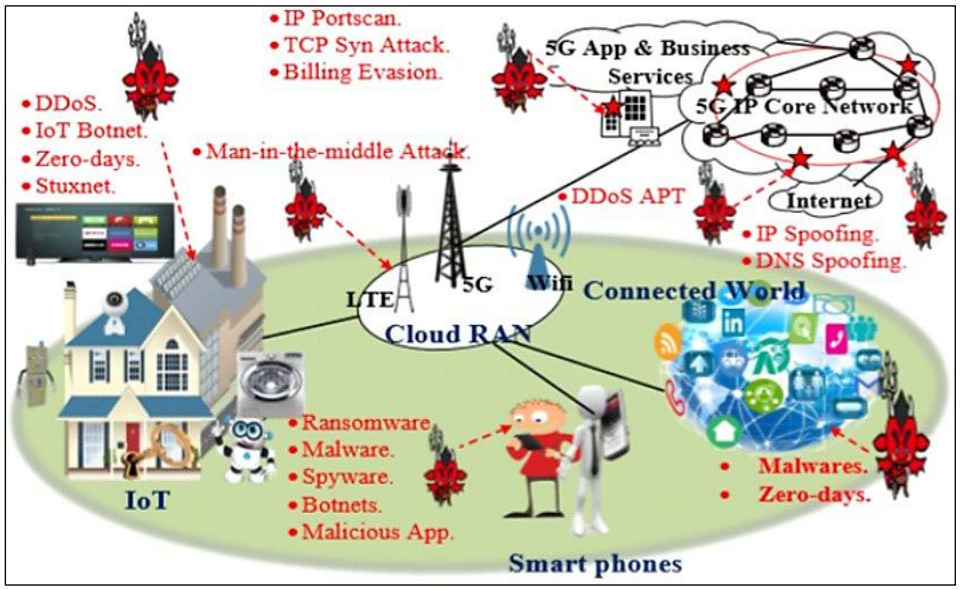
intelligence (AI) in information warfare in the future. It will play an important role in terms of defence, and it is expected that related policies and detailed legislation will need to be prepared.

Table 3. General intelligence operational authority "example of CIA operational authority".

Division	Detail
HUMINT	Acquiring information through local agents and carrying out operations through specific missions
Cyber Operation	Responding to cyber threats in cyber space or performing operations through cyber attacks
Special Operation	Conducting operations to eliminate threats related to national security
Propaganda Operation	Conducting operations to shape public opinion or change the position of the international community and other countries

So far, the operational authority has no judicial authority and is merely a division of the business. In the future, in information warfare where artificial intelligence (AI) is used, it should evolve in the direction of including detailed content and presenting judicial authority due to damage to private facilities caused by artificial intelligence (AI).

Figure 4. An example of operational authority of information warfare using artificial intelligence (AI)[32].



As such, legislation prior to the advent of AI is expected to undergo more than 50% revision and supplementation after the advent of AI, and preparing a judicial basis to support the paradigm of the times is the state's responsibility to maintain national safety and legal order.

As a model for field application, the operation area setting suggests the setting of operational rights in accordance with Article 5 (security area) of the Presidential Security Service (PSS) 'Act on the Protection of the President'. In Korea, for the purpose of protecting the president, a special law-style security area is designated in the security area to temporarily restrict the basic rights of the people. As shown in <Figure 4> above, legislation is needed to designate the area where information warfare is conducted as the operational area of the Intelligence Agency. As for the validity of this model, North Korea's security force, which has been in a military confrontation with South Korea for 70 years, is mobilized to guard the Kim Jong-un family, which accounts for 10% of the entire North Korean army, but the Presidential Security Service has a small scale of less than 700 soldiers. The presidential security mission is being completed with man-

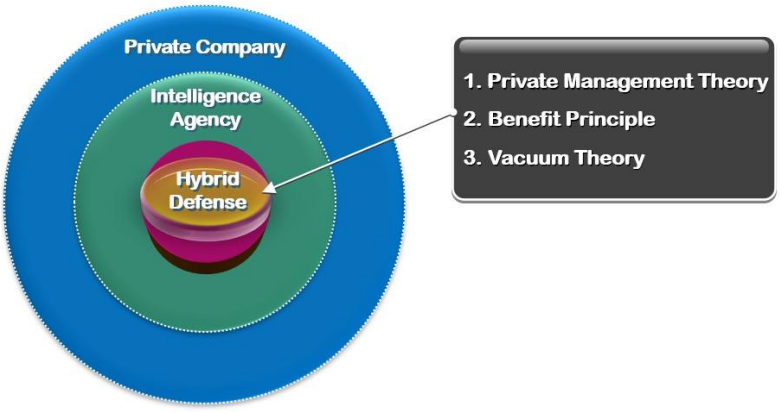
power, and this basis reflects the greater recognition of public interest between national security and the people's discomfort with temporarily controlling the democratic society in free countries.

Also, as shown in <Figure 4>, the US CIA's drone operation model is the most realistic facility operation model for operational control. Drones are operated by the CIA, but facilities such as drones are a dual management system that exists in air bases, aircraft carriers, and embassies. Therefore, facilities required for information warfare occurring in the country are established in division-level military units within the territory of the country to maintain security, and operations are operated by the intelligence department near the military unit that manages the city center, and through an appropriate combination of SIGINT, ELINT, TECHINT and HUMINT It is a COMINT model.

5.2. Privatization

In order to keep up with the rapid development of artificial intelligence (AI) technology, growth through the efficiency of private companies is essential. Discussion on the use of artificial intelligence (AI) in information warfare can present a hybrid information warfare model through private participation and cooperation from the typical perspective of expanding the size of the Intelligence Agency.

Figure 5. Hybrid defense model.



First, if the Private Management Theory is applied, economic feasibility and efficiency have al-ready been proven through private military companies (PMCs) and international intelligence companies, and the effectiveness of operations for continuous profit pursuit can be more sincere than the Intelligence Agency. In this process, the competition of private companies brings about in-novation, and eventually the burden of the Intelligence Agency can be reduced.

Second, if the Benefit Principle is applied, the threat posed by artificial intelligence (AI) increases, but the national budget is limited and users are required to bear some of the services necessary for public safety. Big tech companies and the financial sector that will face intelligence (AI) threats can take a big part. In addition, it will be effective overseas where the influence of the domestic government is less than that of the domestic government.

Third, if the Vacuum Theory is applied, the state and the private sector mutually complement each other to enhance the effectiveness of the vacuum state that was created when the state and the private sector failed to reach each other, and the intelligence department provides a certain part related to artificial intelligence (AI) to the private sector. It is to increase efficiency through the management expertise of private companies while entrusting them to companies and guaranteeing their goals through supervision and regulation.

Finally, these privatization policies are based on the principle of market economy. When the threat of artificial intelligence (AI) increases, the government or corporations spend more to protect assets, and when the economy is activated, more tax revenue and profit generation are threatened. This means that we can secure more budget to respond to. On the other hand, even if the threat of artificial intelligence (AI) increases, if the economy is in a slump, governments and businesses may have limited or unable to respond to the threat.

As such, the threat of artificial intelligence (AI) and the market economy have a close relationship, and the level of defense expands along with the market economy according to the asset size, vulnerability, and threat level of the government and companies, but in socialist dictatorships such as China and North Korea, privatization It has a structural contradiction that cannot be done, so the more artificial intelligence (AI) technology expands, the more it will face limitations.

6. Suggestions

I would like to introduce a case of the dilemma between high-tech machines and human intuition. In 1983, at the height of the Cold War, Lieutenant Colonel Stanislav Petrov, commander of the Soviet Air Defense Force's watch, received an alert from the Serpukhov-15 satellite control center in the Soviet Union that "the United States had launched an ICBM into the Soviet Union." Soon, ICBMs were confirmed to have increased to five, and there was a case in which World War III almost broke out.

At that time, the Soviet Union was using state-of-the-art science and technology to detect and detect the launch of nuclear missiles. It was a time when US President Ronald Reagan had criticized the Soviet Union as an “evil empire” and was ahead of the Able Archer 83 exercise, a preemptive nuclear strike exercise with NATO. It was not surprising that nuclear missiles stationed in Germany and Italy were attacking the Soviet Union, as Soviet leader Yuri Andropot was suffering from chronic illness.

But Stanislav Petrov, on the verge of starting a nuclear war, said, “If the US really starts a nuclear war, it will launch all ICBMs together. But now the computer has caught only five. Therefore, this must be a computer error or an error in judgment by the satellite for detection,” and based on human intuition, he reported to the upper level, “It seems to be a computer error (Кажется, это ошибка компьютера.)”. As a result, the World War III never happened.

This case sheds light on what artificial intelligence (AI) technology is currently replacing the role of humans. The use of artificial intelligence (AI) in information warfare offers efficiency and performance gains, but also reminds us of the importance of human-driven ethical aspects. First, the use of artificial intelligence (AI) can be biased information reflecting prejudice according to learning data. When artificial intelligence (AI) automates missions based on this information, human judgment is eliminated, which can lead to errors in critical decisions and strategies. Second, although certain risks can be predicted through artificial intelligence (AI) predictive modeling, it is necessary to clearly recognize that not all predictions may be accurate, avoid hasty judgments, and take a cautious approach in which human judgment is applied to the decision-making process.

I hope that this study will be published in English to reach more researchers around the world. Also, there are people who were born in North Korea in Northeast Asia and are not guaranteed their basic rights as human beings despite their will and efforts. Similarly, I would like to conclude this study with the desire to think deeply about the fear of an era in which artificial intelligence (AI) controls and monitors humans at some point due to the indiscriminate expansion of artificial intelligence (AI).

7. References

7.1. Journal articles

[1] Min BW. Security Discourse and International Politics -Focusing on Historical Changes in Security Concepts. *Peace Studies*, 20(2), 203-239 (2012).
[2] Cho YY. Uncertainty in Northeast Asian International Relations and Korea. *The Journal of Political Science & Communication*, 19(3), 27-56 (2016).
[3] Cho WD. US-China Competition and Trump Administration's South Asia Strategy: Focusing on India and Pakistan. *National Strategy*, 25(3), 121-148 (2019).
[4] Zhang D & Lei L & Q Ji Q & Kutan AM. Economic policy uncertainty in the US and China and Their Impact on the Global Markets. *Economic Modelling*, 79, 47-56 (2019).
[5] Hyun IT. Hegemonic War between the United States and China and South Korea's Strategy. *New Asia*, 28(2), 57-69 (2021).
[6] Choi JD & Ahn MS. US-China Hegemonic Competition, Neighboring Countries' Response and South Korea's Strategy. *The Journal of Humanities and Social Science (HSS21)*, 12(3), 2981-2994 (2021).
[7] Fajgelbaum PD & Khandelwal AK. Biden JR & Joseph R. The Economic Impacts of the US-China Trade War. *Annual Review of Economics*, 14, 205-228 (2022).
[8] Yun M. Global Strategic Games of the US-West and Russia-China: War and Peace of Global Hegemonic Clashes. *The Journal of Peace Studies*, 23(2), 7-41 (2022).
[9] North Korea's Complex Strategy to Avoid Comprehensive Sanctions. *Korea Research Institute for National Strategy*, 8(2), 27-54 (2023).
[10] Kim YJ. Geopolitical Buffer System Theory and Korea. *National Security and Strategy*, 14(3), 1-46 (2014).
[11] Shifrinson J. Security in Northeast Asia: Structuring a Settlement. *Strategic Studies Quarterly*, 13(2), 23-47 (2019).
[12] Kang JI. Analysis of the Geopolitical Value of the Korean Peninsula and the Military Competition between the U.S. and China. *Military History*, 122, 375-419 (2022).
[13] Yang IB. A Study on the Relationship between Korean Soft Power and Promotion in the Korean Wave: Focusing on the Korean Wave Drama Descendants of the Sun. *Korea and Global Affairs*, 6(2), 73-92 (2022).
[14] Lim SJ & Dong G. Exploring the Global Acceptance of K-Pop and its Direction of Development. *Korea and World Review*, 4(4), 153-177 (2022).
[15] Seol IH. Nontraditional Security Threats and Security Cooperation among South Korea, the U.S, and Japan. *Journal of Korean-Japanese Military and Culture*, 31, 49-79 (2021).
[16] Kang DW. The Current Status of North Koreans' Access to Outside Information and Changes in Their Perception. *Korea and Global Affairs*, 5(2), 97-131 (2021).
[17] Moon HY. Restrictions on Digital Rights in North Korea. *Journal of North Korean Studies*, 47(1), 192-227 (2022).
[18] Shim CS. A Study on the Development of K-drama Contents -Focusing on Chinese Audience's Comments on the Descendant of the Sun-. *The Journal of Image and Cultural Contents*, 10, 45-60 (2016).
[19] Yan Ma Y & Fang X. Development and Characteristics of The Image Fantasy Genre in Korea and China from the 1990s to the 2010s -Focusing on The Gingko Bed, A Terra-Cotta Warrior, Goblin: The Lonely and Great God, Startling by Each Step-. *Asia-pacific Journal of Convergent Research Interchange*, 8(3), 133-144 (2022).
[20] Joo J & Kim B & Chung J. North Korean Refugees' Media Use and Social Capital A Focus on Trust, Network, and Adaptation. *Korean Journal of Journalism & Communication Studies*, 63(4), 45-82 (2019).
[21] Choi CH& Park JB & Kim JG. A Comparative Analysis of Cultural Power as a Soft Power among National Power. *Journal of Digital Convergence*,12(6), 55-68 (2014).

[23] Jo S. A Study on the Creation of the Office of National Security's National Security Investigation Headquarter. *Korean Police Studies Review*, 22(1),241-258 (2023).
[24] Seol IH & Bae HY. The Ukrainian War and Future Warfare: Implications for the Indo-Pacific Region and the Korean Peninsula. *Journal of National Defense Studies*, 66(2), 75-110 (2023).
[28] Kim SR. Legal Tasks and Prospects of the Fourth Industrial Revolution and the AI Era. *Law Review*, 18(2), 21-57 (2018).
[29] Siau K & Wang W. Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. *Journal of Database Management*, 31(2),1-14 (2020).
[30] Kwon SJ. A Study on cyber security in the age of artificial intelligence (AI*). Human Rights Law Review*, 27, 35-74 (2021).
[31] Jo S. The Relationship between Anti-communist Investigation and Domestic Politics. *Korean Security Journal*, 74, 231 -248 (2023).
[32] Kim S & Kim S & Park B & Jeong U & Choo H & Yun J & Kim J. Cyber Electronic Warfare Technologies and Development Directions. *The Journal of Korean Institute of Electromagnetic Engineering and Science*, 32(2), 119-126 (2021).

7.2. Additional References

[22] Owen JM. Balancing Soft and Hard Power: China, Russia, and the United States. Soft Power and the Future of US Foreign Policy. Manchester University Press (2023).
[25] http://news.kmib.co.kr/ (2015).
[26] http://monthly.chosun.com (2011).
[27] https://www.newdaily.co.kr/ (2015).

8. Appendix

8.1. Author's contribution

	Initial name	Contribution
Author	SJ	-Set of concepts ☑ -Design ☑ -Getting results ☑ -Analysis ☑ -Make a significant contribution to collection ☑ -Final approval of the paper ☑ -Corresponding ☑ -Play a decisive role in modification ☑ -Significant contributions to concepts, designs, practices, analysis and interpretation of data ☑ -Participants in Drafting and Revising Papers ☑ -Someone who can explain all aspects of the paper ☑