Asian Journal of Innovation and Policy (2020) 9.1:056-078
DOI: https://doi.org/10.7545/ajip.2020.9.1.056

The Threat of AI and Our Response: The AI Charter of Ethics in South Korea

Ha Hwang[*], Min-Hye Park[**]

Abstract Changes in our lives due to Artificial Intelligence (AI) are currently ongoing, and there is little refutation of the effectiveness of AI. However, there have been active discussions to minimize the side effects of AI and use it responsibly, and publishing the AI Charter of Ethics (AICE) is one result of it. This study examines how our society is responding to threats from AI that may emerge in the future by examining various AIECs in the Republic of Korea. First, we summarize seven AI threats and classify these into three categories: AI's value judgment, malicious use of AI, and human alienation. Second, from Korea's seven AICEs, we draw fourteen topics based on three categories: protection of social values, AI control, and fostering digital citizenship. Finally, we review them based on the seven AI threats to evaluate any gaps between the threats and our responses. The analysis indicates that Korea has not yet been able to properly respond to the threat of AI's usurpation of human occupations (jobs). In addition, although Korea's AICEs present appropriate responses to lethal AI weapons, these provisions will be difficult to realize because the competition for AI weapons among military powers is intensifying.

Keywords Artificial Intelligence, AI threats, AI Charter of Ethics

I. Introduction

The development of artificial intelligence (AI) is expected to bring about a radical change in our society. Those who lead and advocate the development of AI claim that AI will bring to humanity a level of convenience and happiness previously unimaginable. Medium, the online social journalism platform in the United States, summarizes the eight advantages of AI as follows: 1) Reduction in human error; 2) Takes risks instead of humans; 3) Available 24/7 (AI works full time without break periods); 4) Helps in repetitive jobs; 5) Digital assistance;

Submitted, April 15, 2020; 1st Revised, April 28, 2020; Accepted, April 29, 2020
* Main author and corresponding, Associate Research Fellow, Korea Institute of Public Administration, Seoul, Korea; hahwang@kipa.re.kr
** Co-author, Commissioned Researcher, Korea Institute of Public Administration, Seoul, Korea; rabomh@unist.ac.kr

6) Faster decisions; 7) Daily applications (e.g. Apple's Siri, Google's OK Google); and 8) New inventions (Kumar, 2019).

Changes in our lives due to AI are currently ongoing. IPsoft, an American technology company, introduced Amelia, an AI digital assistant in 2014. Amelia specializes in human resource management systems, IT service management, banking, healthcare, insurance, retail, and other services. Through years of continuous development, Amelia is now an AI employee used by many global companies such as Shell Oil and Allstate Insurance ("The World's First Marketplace for Digital Employees," n.d.). Another example is bomb detection, where AI is applied to analyze the number and location of unexploded bombs based on the number of bombs dropped and the number of craters left when the bombs exploded. This method's detection accuracy is greater than 160% compared to other technology (Ohio State University, 2020).

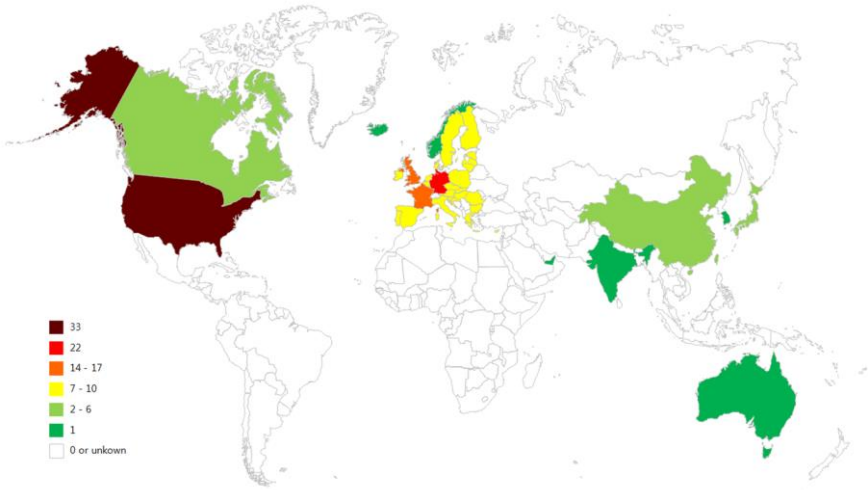
There are few arguments that refute the effectiveness of AI. However, the problem is that the changes caused by AI will not be limited to the levels of additional utility in the current state; the ripple effect is expected to affect not only economic and social systems, but also international politics and value systems. A report published by the World Economic Forum ("The Global Risks Report 2017," 2017) provides valuable insights regarding this issue. According to the report, AI and robotics technologies are expected to provide both the most benefits to humanity and the most side effects. In addition, of the twelve key emerging technologies, AI and robotics are designated as the technology combination most likely to exacerbate global risks in economic, societal, geopolitical, and technological sectors; understandably, it is ranked first among technologies that require better governance.

Recently, there have been active discussions to minimize the side effects of AI and to apply it responsibly. Livingston and Risse (2019) warn that AI technology should be developed in a way that ensures human rights; otherwise, the time will come when the boundaries between humans and AI will collapse. The "Portrait of Edmond Belamy", the first artwork created by AI, was sold for $432,500 by the global auction house Christie's on October 25, 2018. The auction price was approximately 45 times higher than the expected price. The AI was trained using 15,000 works of art created from the 14th to the 20th centuries using the Generative Adversarial Network (GAN) method ("Is artificial intelligence set to become art's next medium?," 2018). Although the first AI-created artwork was sold at such a high price, whether it can be called, as *art* is still questionable. To date, AI-generated art cannot completely replace human art. However, art can be redefined, and even in the realm of creative activities, the possibility of a society in which AI replaces humans cannot be ruled out (Wang, 2019).

The international move to set standards for the proper use of AI is being driven by the European Union. The European Commission (EC) announced the White

Paper on Artificial Intelligence and a European Strategy for Data on February 19, 2020 in Brussels as part of its strategy ("Shaping Europe's digital future – Questions and Answers," 2020). This report includes a strategy-setting plan to simultaneously achieve two goals: securing the AI ecosystem of excellence for economic growth, and the AI ecosystem of trust for people-oriented AI development. In particular, the White Paper mentions possible challenges as well as strategies to maximize the benefits of AI.

Another effort to use AI for proper purposes and to prevent side effects is the AI Charter of Ethics (AICE). To date, more than one hundred AICEs have been published worldwide ("Al Ethics Guidelines Global Inventory," n.d.; Jobin, Ienca, & Vayena, 2019). European countries have published fifty-eight AICEs, the largest number by a continent. The United States and Canada come next with thirty-nine AICEs. Asian countries have published eighteen AICEs. Middle Eastern countries and Oceania countries have each published one AICE, based on our research. Comparing countries, the United States has published the largest number of AICEs, followed by Germany, Britain, and France (see Figure 1). Efforts to create new and better AICEs worldwide continue. An example is the Responsible AI in Africa Network program ("KNUST AI NETWORK," n.d.) launched in March 2020. A collaboration team representing Kwame Nkrumah University of Science and Technology in Ghana and the Institute for Ethics in Artificial Intelligence is leading the project, and expects to publish an African AICE in the near future ("Responsible AI in Africa Network," n.d.).



Source: The authors, based on UNCTAD data (2017) and World Bank World Development Indicators (2017).

Note 1: From the two data sources, AICEs published by countries or the EU are counted. Excluding duplicates, a total of 117 AICEs are included in the figure. Seven AICEs were

published by the EU, so all 27 EU members show seven and above counts for this reason.

Note 2: The map is possibly biased toward English using countries and European countries due to limitations in the authors' linguistic capabilities and greater availability of English data.

Note 3: We found three AICEs published by Singapore, but Singapore does not appear on the map owing to the very small size of its territory.

Note 4: For South Korea, only the Kakao Algorithm Ethics Charter was included because of the reasons mentioned in Note 2.

Figure 1 Number of AI Charter of Ethics documents published by country

We studied how our society is responding to the threat of AI by examining the AIECs published in South Korea. This study is organized as follows. Section 2 summarizes the future threats expected from AI. The latest news articles, related websites, reports by governments and companies, and research papers were investigated and were classified into seven threats representing three primary categories. Section 3 introduces seven AICEs in Korea and categorizes their contents into thirteen topics. Section 4 examines how well the Korean society is responding by comparing the future threats that AI may cause with the contents of Korea's AICEs. Along with a brief summary, the direction to properly respond to future threats arising from use of AI is discussed in the final section.

II. Expected Threats by Artificial Intelligence

The threats arising from the cross-social application of Artificial Intelligence (AI) vary. We searched Internet news articles, recent research reports, and academic studies discussing such threats to synthesize them. The survey outline of the research materials is shown in Table 1. After reviewing the collected data, we identified seven expected threats and classified these into three categories: 1) AI's value judgment, 2) malicious use of AI, and 3) human alienation (see Table 2).

Table 1 Survey outline of potential threats by AI

	Internet news articles	Research reports and academic studies
Issued years	2016 – present	1976, 2007, 2016 – present
Search keywords	Artificial Intelligence, AI threats, AI ethics, AI harm, AI side effects, AI social problem, AI dark side, malicious use of AI	
Search platform	Google	Google scholar, Google

Table 2 Seven expected threats by AI

Category	Expected threats of AI
AI's value judgment	1. Human discrimination in AI 2. AI's weighing of human value
Malicious use of AI	3. Lethal AI weapons 4. AI-based cyber attacks 5. Excessive privacy intrusion
Human alienation	6. AI's usurpation of human occupations 7. Deepening the alienation of the digitally vulnerable

1. AI's Value Judgment

As AI becomes more widely used, situations may arise in which the AI judges moral values. Training using more value judgment cases will advance AI further, and key decisions may be made by AI instead of humans in the future. This may lead to AI judgment problems.

1.1 Human discrimination in AI

AI generates decision algorithms through learning based on human data. Therefore, AI can commit the same errors as humans, and these errors can be reinforced through self-learning. Human discrimination in AI refers to a situation where AI discriminates against a specific group because of such errors. The AI recruitment system used by the American E-commerce company Amazon.com is an example. The company organized a team to develop an AI system for recruitment in 2014 based on Amazon's recruiting data from the past 10 years. However, in 2015, they found that their system was discriminating against female candidates. The system penalized the resumes that included words such as "women" and placed high value on words such as "executed" or "captured," which were usually found in male candidates' resumes. Amazon attempted to modify the AI program, but the company finally stopped the project and disbanded the team in 2018 because it was difficult to predict what other types of discrimination could be generated by the AI in the future (Dastin, 2018). Another example is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system used by the U.S. states of New York, Wisconsin, California, Florida's Broward County, and other jurisdictions (Kirkpatrick, 2017). The AI helps to evaluate the likelihood that a prisoner will commit a crime again and helps to determine the probation guidelines for the prisoner. This AI does not use skin color as input data when learning, but it nevertheless tended to discriminate against black people by judging their recidivism rate to be almost twice as high as white people (Angwin, Larson,

Mattu, & Kirchner, 2016). A larger problem is that only 61% of potential repeat offenders predicted by COMPAS reoffended within two years (versus the probability of random selection being 50%, 61% is a little higher accuracy) (Zuiderveen Borgesius, 2018). These are only a few cases in which human discrimination in AI has been revealed, indicating the possibility that AI can be trained to intentionally discriminate against different humans.

1.2 AI's weighing of human value

AI's weighing of human value refers to a situation in which AI judges the value of different people and ranks these values. This situation can have serious social consequences if it leads to a decision involving the death of an undervalued life. This includes the Trolley Dilemma; whether to make a decision to save the majority by sacrificing the minority in a situation where either sacrifice is inevitable (Thomson, 1976). It is expected that AI will face this problem at some point in time, which is a problem to which even humans cannot provide a clear answer. In South Korea, the release of a level 3 autonomous vehicle on dedicated lanes will be available from July, 2020. As a result, it is expected that various AI-related social conflicts will be experienced in the transportation sector first. To prepare for this, the Korean government enacted the "Act on Promotion and Support for the Commercialization of Autonomous Vehicles," which is scheduled to take effect on May 1, 2020 (*Korea Law Information Center*, n.d.). In addition, the Ministry of Land, Infrastructure and Transport is making efforts to establish safety standards for level 3 autonomous vehicles for the first time in the world and to prepare guidelines for ethics applicable to autonomous vehicles ("Collect Opinions on Self-driving Ethics Guidelines," n.d.).

2. Malicious Use of AI

Use of AI for malicious purposes can lead to dangers far beyond current expectations. Twenty-six AI experts held a workshop entitled 'Bad Actor Risks in Artificial Intelligence' in February 2017, in Oxford, UK. It was jointly held by the Future of Humanity Institute and the Centre for the Study of Existential Risk, and a report of their findings, entitled "The malicious use of artificial intelligence" was released in February, 2018 (Brundage et al., 2018). The report identifies three different security threats that can arise from the malicious use of artificial intelligence: digital security, physical security, and political security. Similarly, in this study, we classified threats from malicious use of AI as lethal AI weapons, AI-based cyber-attacks, and excessive privacy intrusion.

2.1 Lethal AI weapons

When AI is incorporated into a lethal weapon, it can attack people who are considered enemies without direct control. These weapons may inflict casualties on innocent people if the AI fails. Use of AI can also generate a gap in military power between countries with and without such technology and can be used for terrorist attacks targeting key national figures or mass killings against an unspecified majority. According to Human Rights Watch ("Killer Robots," n.d.), an international non-governmental organization (NGO) in New York, autonomous weapons (also known as killer robots) are being developed and deployed in China, Israel, Russia, the United Kingdom, the United States, and South Korea. In South Korea, the unmanned SGR-A1 system, which belongs to the Lethal Autonomous Weapons Systems (LAWS) category, has been developed for national defense on Demilitarized Zone (DMZ) and unveiled in 2007 (Kumagai, 2007). There are pros and cons regarding the use of LAWS. From the pros viewpoint, LAWS represent various military advantages, as they do not require communication links, fewer soldiers are needed, and because they are not humans, war crimes such as rape and indiscriminate killing can be reduced. However, from the cons viewpoint, there exist concerns regarding malfunction and ethical matters, as these machines can decide to kill human beings without oversight (Surber, 2018).

2.2 AI cyber-attacks

AI cyber-attacks can be subdivided into two categories. One is attacks on AIs; hacking an AI that controls a physical system and damaging it. For example, hacking into a self-driving vehicle control system could cause serious social confusion and damages if it interferes with the proper operation of the control system and vehicles (Brundage et al., 2018). Another example is the smart Cyber-Physical System (sCPS). Smart governance, smart buildings, and smart transportation are well-known examples of a Cyber-Physical System (CPS). Recently, with the advent of technologies, such as the Internet of Things (IoT), AI, wireless sensor networks (WSNs), and Cloud Computing, smart "anything" has become possible, and the concept of smart CPS (sCPS) has emerged (Kaloudi & Li, 2020). As our hyper-connected society progresses, a malicious cyber-attack on an AI that controls sCPS does not end with the failure of that system, and the possibility of an entire society being paralyzed by cascading failures of connected systems cannot be excluded.

The other category is to attack using AI; AI-based cyber-attacks such as malware, socialbots, voice synthesis, and other attacks (Kubovic, Kosinar, & Janosik, 2018). For example, DeepLocker, a highly advanced malware, applies AI trained through Deep Neural Network algorithm to conceal itself until reaching its target and then usurps operations after reaching the target. Socialbots are an example of AI designed to communicate with people on the

Internet with the purpose of changing people's political orientation or views on specific issues. Speech synthesis involves an AI that learns and reproduces the voice and tone of a specific person. There have been cases in which a politician's voice file or video has been faked and distributed on the Internet, or in which artificial intelligence has been used to commit financial fraud through telephone calls (Kaloudi & Li, 2020).

2.3 Excessive privacy intrusion

Excessive privacy intrusion is the case in which information is used as data to train an AI and/or when information predicted through an AI violates human privacy. This threat can be divided into three categories according to the type of information collected for and predicted by AI: personal information, consumption behavioral information, and emotional information. First, collecting patient medical data to improve the performance of medical AI can compromise patient privacy (Olson, 2018). This is a problem that can occur in various fields other than medical practice. In response to these problems, methods such as filtering sensitive personal information are being evaluated (Rohringer, Budhkar, & Rudzicz, 2019).

Second, collecting consumers' behavioral information to predict their product orders in advance falls under the consumption behavioral information category. With the increase in online retailers, the competition for rapid delivery is fierce. AI's capability to predict purchasing behavior based on consumers' purchasing information is a powerful tool for gaining an edge in this competitive environment. This type of change not only provides enormous convenience to customers, but also leads to the transformation from a 'shopping-then-shipping' to 'shipping-then-shopping' business model (Davenport, Guha, Grewal, & Bressgott, 2020). However, it should be noted that the more people's behavioral information is exposed and the more accurately it is predicted, the greater the scrutiny and associated privacy risks.

Lastly, attempts are being made to develop AI that reads emotions by analyzing human expressions. While some companies have used this technique to save time when recruiting for jobs, some experts question the accuracy of AI owing to the complex correlations between human expressions and emotions (Devlin, 2020). Furthermore, excessive invasion of privacy through AI can bring about various threats that we have never experienced, such as obtaining a password using a person's heart rate or body temperature data (Sedenberg & Chuang, 2017).

3. Human Alienation

3.1 AI's usurpation of human occupations

The problem with human occupations being replaced by AI is the most immediate problem, and many studies have been published, but no clear solution has been proposed. The more jobs that AI can take over from humans, the fewer jobs people will be able to perform. Food order kiosks that replace personal service and autonomous trucks that replace truck drivers are already commercially available. Some experts believe that some jobs will disappear owing to AI, but the jobs will increase overall as new jobs appear in the fields of technology and science ("How will AI change the future of work?," n.d.). The dilemma, however, is that the majority of those in the occupations that are expected to be replaced first are socio-economically disadvantaged, and it will be difficult to re-educate them for employment in highly technical fields.

3.2 Exclusion of digital vulnerabilities

When AI is widely used in everyday life, it is possible that people who have limited access to digital devices, such as the elderly, are marginalized. According to one survey (*A Survey on the Actual Condition of Digital Difference*, 2019) conducted by the National Information Society Agency (NIA) in Korea, it is expected that socially marginalized people are more likely to be further marginalized owing to the development of AI. Regarding the level of digital information competency, which indicates the ability to use PCs and mobile devices, the competency of the elderly was approximately 50% of the country's average. As advanced countries with active AI research are rapidly increasing in elderly populations, the problem of alienation of the digitally vulnerable is expected to become more serious in the future as AI becomes more commonplace.

III. AI Charter of Ethics in Korea

1. Objectives of AI Charter of Ethics

The AI Charter of Ethics (AICE) was created to develop and utilize AI in a manner that ensures human safety and improves well-being by accurately identifying the positive and negative impacts that may arise from AI. In Korea, there are seven documents that can be classified as AICEs. These are:

- Draft of the Robot Ethics Charter (DREC) (Ministry of Commerce Industry and Energy, 2007, March);

- Kakao Algorithm Ethics Charter (KAEC) (Kakao Corporation, 2018, January);
- Ethical Guidelines for Intelligence Information Society (EGIIS) (Ministry of Science and ICT & National Information Society Agency, 2018, June);
- Intelligent Government Ethics Guideline for Utilizing Artificial Intelligence (IGEG) (Korean Internet Ethics Association & National Information Society Agency, 2018, December);
- Charter of Artificial Intelligence Ethics (CAIE) (Korea Artificial Intelligence Ethics Association, 2019, October);
- Principles for User-Oriented Intelligence Society (PUOES) (Korea Communications Commission & Korea Information Society Development Institute, 2019, November); and
- Ethical Guidelines for Self-driving Cars (EGSC) (Ministry of Land Infrastructure and Transport & Korea Agency for Infrastructure Technology Advancement, 2019, December).

These documents commonly state that the AICE is essential to ensure human safety and prevent misuse of AI. For example, creating a human-centered (Ministry of Science and ICT & National Information Society Agency, 2018, June) and safe AI service environment from the technical and social risks of AI (Korea Communications Commission & Korea Information Society Development Institute, 2019, November), such as killing humans by robots equipped with AI (Ministry of Commerce Industry and Energy, 2007, March), is one of the important objectives of an AICE. In addition, such documents aim to prevent the misuse of AI (Ministry of Science and ICT & National Information Society Agency, 2018, June) and to pursue the benefits toward the happiness of humanity (Kakao Corporation, 2018, January; Korea Artificial Intelligence Ethics Association, 2019, October); that is, to utilize AI for the public good (Ministry of Land Infrastructure and Transport & Korea Agency for Infrastructure Technology Advancement, 2019, December). Table 3 summarizes the objectives of the seven AICEs published in Korea with brief overview.

Table 3 Overview of the seven AI Charters of Ethics in Korea

Name	Date	Institution	Composition	Purpose
Draft of the Robot Ethics Charter (DREC)	Mar. 2007	Ministry of Commerce, Industry and Energy	7 chapters	To identify human-centered ethical codes for the coexistence of humans and robots
Kakao Algorithm Ethics Charter (KAEC)	Jan. 2018	Kakao Corporation	6 chapters	To apply social ethics to all efforts related to algorithms To seek benefits and happiness for humankind
Ethical Guidelines for Intelligence Information Society (EGIIS)	Jun. 2018	Ministry of Science and ICT & National Information Society Agency	6 chapters	To realize the value of sustainable symbiosis To move toward a safe and reliable intelligent information society
Intelligent Government Ethics Guideline for Utilizing Artificial Intelligence (IGEG)	Dec. 2018	Korean Internet Ethics Association & National Information Society Agency	10 chapters with 20 articles	To respond to the problems that can be caused by AI-using government services according to the basic plans of 'intelligent government' announced in March, 2017
Charter of Artificial Intelligence Ethics (CAIE)	Oct. 2019	Korea Artificial Intelligence Ethics Association	5 chapters with 37 articles	To recognize the adverse effects and the risks of AI and find ways to respond to them
Principles for User-Oriented Intelligence Society (PUOES)	Nov. 2019	Korea Communications Commission & Korea Information Society Development Institute	7 items	To suggest public rules for a safe intelligent information society protected from the risks that can be caused by the adoption of new technology
Ethical Guidelines for Self-driving Cars (EGSC)	Dec.2019	Ministry of Land, Infrastructure and Transport & Korea Agency for Infrastructure Technology Advancement	6 chapters with 29 articles	To improve human safety and welfare To ensure safe and convenient freedom on right of mobility To consider human life first before animal lives or property damage To minimize personal and social losses from accidents

2. Contents of AI Charters of Ethics

Content analysis was conducted on the seven Korean AICEs. As a result, they were divided into three categories: protection of social values, AI control, and fostering digital citizenship; they were further divided into fourteen sub-categories. The contents of the seven AICEs classified into fourteen sub-categories are listed in the Appendix.

2.1 Protection of social values

The protection of social values consists of four sub-categories as follows. First, prevention of social discrimination is considered. This states that AI training data should not be biased and should not discriminate based on gender, age, race, disability, etc.

Second, inclusion of society as a whole must be considered. As per this, by using AI, all members of society should be able to enjoy the resulting benefits, and to achieve this, discriminatory factors should be excluded at all stages of AI development and implementation.

Third, human dignity should be respected. This states that the development of AI should be accomplished without threatening human dignity.

Lastly, humankind's benefits and happiness should be pursued. This states that the use of AI should ultimately follow the direction that benefits humanity, that is, to help humans pursue personal happiness.

2.2 AI control

The AI control category consists of eight sub-categories as follows. First, an explainable AI algorithm is necessary. This means that it should be possible to fully disclose and explain to the user as well as the general public the type of algorithm applied to AI. In particular, companies such as Kakao should also consider maintaining technology security so as not to undermine their competitiveness. Therefore, in the case of AI developed by private companies, debates regarding to what extent an algorithm should be explained are expected.

Second, the use of data based on social ethics must be considered. Data for AI training must be legally collected with the consent of the owners, and protection must be provided for personal information and other data that has not agreed to be released. EGSC (2019, December) contains more specific content regarding information ethics and security, which can be considered precautionary measures as level 3 self-driving cars become commercially available in Korea starting from July 2020.

Third, malfunctions of AI and consequent risk situations must be considered and prepared for. It is necessary to prepare not only for malfunctions of AI but also for situations caused by deliberate AI hacking. In particular, CAIE (2019,

October) states that an AI kill switch is an essential element, because AI can learn on its own, and not necessarily toward an intended goal.

Fourth, clear division of responsibilities is necessary. In the future, AI-applied products will become commonplace, leading to legal disputes that may arise from AI malfunctions or unexpected problems. The legal responsibility for this type of situation should be clearly clarified by each agent. In particular, EGSC (2019, December) focuses more specifically on the division of responsibilities in the event of an autonomous vehicle accident in preparation for the commercialization of level 3 self-driving cars.

Fifth, ultimately, humans should be able to control AI. When using AI-applied products, an important issue that arises along with the responsibility is the decision-making issue. This means that human decision-making power should take precedence over the judgment of an AI. AI must maintain its role as a tool to support human choices, and human choices should not be submissive to an AI.

Sixth, limiting the purpose of using AI is important. The purpose of using AI should be limited so that it cannot be used to injure or kill humans. Furthermore, the purpose of AI should be toward maximizing human convenience and improving environmental sustainability.

Seventh, activating post-management systems regarding AI is necessary. The products applying AI require continuous monitoring by sellers, developers, and managers for defects or risks. Because products with AI can learn on their own, it is imperative to assume their possible evolution in unintended directions. From this perspective, the existence of a thorough follow-up management system seems essential.

Lastly, it should be possible to check whether AI is being applied, meaning that consumers should be able to check whether or not AI is applied to products or services before using such. To properly apply the guideline to the general population, consideration should be given to minority groups such as the disabled and the elderly using various notation methods such as voice guidance and easy-to-understand instructions.

2.3 Fostering digital citizenship

Fostering digital citizenship is divided into AI governance and individual competency enhancement. First, it addresses the formation of a culture based on continuous multicultural communication; it is necessary to foster a culture that solves problems arising from AI through continuous multilateral communication throughout the development, production, consumption, and post-management process of AI.

Next, enhancing an individual's capability for using AI is important. AI-applied products and services are expected to become universal. It is necessary to improve the digital capabilities of those who use them, of course, the AI

product itself needs to be offered in a form that can be easily used by anyone. To do this, it is necessary to establish an educational infrastructure at the national or regional level.

IV. Gaps between Threats and Responses

We examined how well the AICEs in Korea respond to the seven threats that may arise from AI. To this end, the thirteen topics contained in the Korean AICEs were reviewed based on the seven AI threats. Table 4 shows the results when matching them. An "O" indicates the case where an appropriate and necessary response to the threat is included in the content, and "X" represents the case where the appropriate response is needed, but there is no relevant content in the AICEs.

First, Korea's AICEs had the most inadequate response to the threat of 'AI's usurpation of human occupations.' The AICEs should have addressed this issue in at least two areas. The first is 'respect for human dignity'. That is, any change arising from AI should not threaten human dignity. The problem involving job replacement by AI has already begun in many countries, but measures for the 'right to live' of those who have lost their jobs have not even begun to be discussed. Without a choice but to do so, modern nations based on liberal democracy and the market economy have neither a way to prevent capitalists from trying to introduce AI to increase productivity, nor have the capability to account for the resulting unemployed population. The second is 'pursuit of human benefit and happiness'. The important point here is "whose" benefits and happiness to pursue are involved. In general terms, it can be interpreted that AI should be used for the benefit and happiness of "everyone," that is, for the *public good*. The meaning becomes clearer when interpreted together with the concept of 'social inclusion as a whole'. From this point of view, someone's "benefit" arising from AI should not bring "misfortune" to others.

Regarding other threats, it appears that Korea's AICEs are responding well to some extent. However, they are more focused in terms of "how to respond", and there is a lack of perspective on "why" such responses are needed. For example, the principles of "how" to respond to the threat of 'human discrimination in AI' were proposed in the content pertaining to 'prevention of social discrimination' and 'inclusion of society as a whole.' On top of this, it is necessary to deal with the reasons for such a response from the perspectives of 'respecting human dignity' and 'pursuing human benefits and happiness.' If AICEs contain only "how", they only provide a recommendation, but if they aim to protect the basic rights guaranteed by the Liberal Democratic Constitution, then legal binding can be applied. In the same vein, the issue of preventing malicious use of AI and

human alienation also requires direct reference to ‘respect for human dignity’ and ‘pursuit of human benefit and happiness.’

We found that the AICEs had fairly an appropriate response to the threat of ‘lethal AI weapons.’ However, ironically, South Korea has been developing a “killer AI robot” for security in DMZ since 2006. This may be because the AICEs are not compulsory, but it is expected that the provisions involving lethal AI will be difficult to implement because there is already competition for the development of AI weapons among various international military powers. Finally, we found that some of the AICEs’ content was not directly related to the threats posed. These are ‘activating the post-management system,’ ‘clear division of responsibility,’ and ‘culture of continuous multilateral communication,’ all of which are related to AI governance in common. Although they do not correspond to specific threats, they are essential for building a social infrastructure tailored to an AI society that can correctly utilize AI and respond quickly and appropriately to future threats.

Table 4 AI Charters of Ethics in Korea and their response to the threats from AI

Contents of AICE \ Expected Threats of AI		AI's value judgment		Malicious use of AI		Human alienation		
		1. Human discrimination in AI	2. AI's weighing of human value	3. Lethal AI weapons	4. AI-based cyber attacks	5. Excessive privacy intrusion	6. AI's usurpation of human occupations	7. Deepening the alienation of the digitally vulnerable
Protection of social values	Prevention of social discrimination	O	O					
	Social inclusion as a whole	O	X					O
	Respect for human dignity	X	O	O	X	X	X	X
	Pursuit of human benefit and happiness	X		O	X	X	X	X
AI control	Explainable algorithm	X				O		X
	Use of data based on social ethics	O				O		
	Prepare for malfunctions and hazardous situations				O			
	Ultimately human controlled				O			
	Limiting the purpose of using AI			O	O	X		
	Activating the post-management system							
	Clear division of responsibility							
	Possible to check whether AI is applied							
Fostering digital citizenship	Culture of continuous multilateral communication							
	Enhancement of AI utilization capabilities						X	O

V. Summary and Discussion

We are currently enjoying various benefits of AI at hand and display the desire to continue to develop new benefits through AI. However, it should be noted that there are side effects that can threaten our society. In this study, the following three threats that may arise from AI were examined: AI's value judgment, malicious use of AI, and AI's usurpation of human occupations. To evaluate whether Korea is publicizing appropriate responses to these threats, we examined seven AICEs published in Korea. The results revealed that the Korean society is not prepared to respond to the threat of 'AI's usurpation of human occupations' at all. In addition, we found that the threat of 'lethal AI weapons' was well covered in the AICEs, but the reality indicates differently.

Some would argue that artificial intelligence's job replacement is just a process on the path to a new society and therefore makes sense. Historically, human society has developed along with technological advances, such that many jobs have disappeared, new jobs have been created, and the social and economic impacts of the process have been absorbed by society as a whole. However, it is doubtful that human society can naturally absorb the socio-economic impacts caused by the fourth industrial revolution represented by the AI revolution and find a new equilibrium. When considering resilience, two components emerge: the force applied and the characteristics of a system. It is difficult to guess how large the socio-economic impact caused by AI will be, but it will be greater than any impact in the past. In addition, humanity living in modern society enjoys more freedom and abundance than ever before. Given to what extent these conditions may change, it is difficult to be optimistic about the future AI society.

Others may argue that the threat of AI job replacement is not mentioned in the AICEs because it is not an ethical issue. This argument makes sense to some and not to others. Some may accuse the capitalist system of replacing one hundred jobs with a single AI as being immoral, but others believe this to be a reasonable choice. However, according to the 'Artificial Intelligence: the global landscape of ethics guidelines' (Jobin et al., 2019), which surveyed 84 AICEs published worldwide, the 'Justice & fairness' and 'Sustainability' categories emphasize the changes in the job market that AI will bring and warns that unemployment due to AI is not expected to occur in a specific region or industry, but rather across society as a whole. In other words, this could be a problem that could threaten the free market economy and democracy of many countries across the world.

Finally, the threat of 'lethal AI weapons' is not an issue that can be addressed at the national level, as it is related to the international arms race. Globally, the controversy over AI weapons is ongoing ("Campaign to Stop Killer Robots," 2020; Gayle, 2019). Even if an international treaty is signed, it will not be easy to monitor its implementation, as AI weapons are not as easily detectable as for

example nuclear weapons, and AI weapons can be produced in various sizes and shapes. Furthermore, even if an AI weapon mounted on a drone were found, it could prove difficult to identify the drone's owner.

There are at least four AI ethics problems currently ongoing in Korea. These are autonomous vehicles accident liability problem, cognitive problems when interacting with AI, AI's personal data collection problem, and AI's job replacement problem. The AICEs in Korea respond to the first three problems with "clear division of responsibility," "possible to check whether AI is applied," and "use of data based on social ethics," respectively. On the other hand, preparations for responding to the last problem are lacking. South Korea is one of the leading countries in the fourth industrial revolution and its industrial structure is highly advanced. For these reasons, it is expected that the Korean society will soon experience the problem of "AI's usurpation of human occupation." Therefore, in order to effectively respond to the threat, discussions on social solutions to this issue should begin immediately.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education [2018R1D1A1B07047033].

Appendix Related provisions by content classification of Korea's seven AI Charter of Ethics documents

Category	Subcategory	DREC	KAEC	EGIIS	IGEG	CAIE	PUOES	EGSC
Protection of social values	Prevention of social discrimination		Chapter 2	Chapter 4	Chapter 8 Article 1 & 2	Chapter 2 Article 13	No discrimination	Article 1.1
	Social inclusion as a whole		Chapter 6	Chapter 2		Chapter 5 Article 36 & 37	Participation	Article 1.3
	Respect for human dignity	Chapter 2, Chapter 5		Chapter 1		Chapter 1 Article 2	Providing people-oriented services	Article 1.1
	Pursuit of human benefit and happiness		Chapter 1	Chapter 1		Chapter 1 Article 1	Providing people-oriented services	Article 1.2
AI control	Explainable algorithm		Chapter 5	Chapter 4	Chapter 6 Article 1, Chapter 9 Article 1	Chapter 2 Article 11	Transparency and explainability	Article 4.4
	Use of data based on social ethics	Chapter 2, Chapter 5	Chapter 3		Chapter 7 Article 1 & 2	Chapter 2 Article 13 & 14, Chapter 3 Article 30	Privacy and data governance	Article 2.5, Article 3.1, Article 6.5
	Prepare for malfunctions and hazardous situations		Chapter 4	Chapter 3	Chapter 10 Article 1 & 2	Chapter 2 Article 12 & 16	Safety	Article 3.3-3.6

	Ultimately human controlled	Chapter 4			Chapter 5 Article 1	Chapter 1 Article 4 & 6 & 7		Article 1.1
	Limiting the purpose of using AI	Chapter 4			Chapter 1 Article 1	Chapter 2 Article 8, Chapter 3 Article 22, Chapter 4 Article 33		
	Activating the post-management system				Chapter 2 Article 2, Chapter 9 Article 2	Chapter 2 Article 19		
	Clear division of responsibility	Chapter 7			Chapter 1 Article 2, Chapter 4 Article 1 & 2	Chapter 3 Article 21 & 23, Chapter 5 Article 35	Responsibility	Article 2.1 & 2.3, Article 4.3, Article 5.1 & 5.4, Article 6.1 & 6.3
	Possible to check whether AI is applied				Chapter 6 Article 2	Chapter 3 Article 29		
Fostering digital citizenship	Culture of continuous multilateral communication			Chapter 5		Chapter 3 Article 31	Participation, Privacy and data governance	
	Enhancement of AI utilization capabilities			Chapter 6				Article 2.4, Article 5.2

References

ALGORITHM WATCH. (n.d.). AI Ethics Guidelines Global Inventory Retrieved April 9, 2020 from https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *PRO PUBLICA*. Retrieved April 8, 2020 from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

A Survey on the Actual Condition of Digital Difference. (2019). *Korean Statistical Information Service*. Retrieved from http://kosis.kr/statisticsList/statisticsListIndex.do?menuId=M_01_01&vwcd=MT_ZTITLE&parmTabId=M_01_01&parentId=I.1;I2.2;110_12017.3;#110_12017.3 (Korean).

Brundage, Miles, Avin, Shahar, Clark, Jack, … Dario. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Retrieved from https://arxiv.org/abs/1802.07228v1.

Campaign to Stop Killer Robots. (2020). Campaign to Stop Killer Robots. Retrieved April 14, 2020 from https://www.stopkillerrobots.org/.

CHRISTIE'S. (2018). Is artificial intelligence set to become art's next medium? Retrieved April 8, 2020 from https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx?sc_lang=en.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *REUTERS*. Retrieved April 1, 2020 from https://uk.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUKKCN1MK08G.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*(1), 24-42.

Devlin, H. (2020). AI systems claiming to 'read' emotions pose discrimination risks. *The Guardian*. Retrieved April 9, 2020 from https://www.theguardian.com/technology/2020/feb/16/ai-systems-claiming-to-read-emotions-pose-discrimination-risks.

DIGITAL Workforce.ai. (n.d.). The World's First Marketplace for Digital Employees. Retrieved April 7, 2020 from https://hire.digitalworkforce.ai/1store/user/home.

European Commission. (2020). Shaping Europe's digital future - Questions and Answers. Retrieved April 7, 2020 from https://ec.europa.eu/digital-single-market/en/news/shaping-europes-digital-future-questions-and-answers.

Gayle, D. (2019). UK, US and Russia among those opposing killer robot ban. The Guardian. Retrieved April 14, 2020 from https://www.theguardian.com/science/2019/mar/29/uk-us-russia-opposing-killer-robot-ban-un-ai.

The Global Risks Report 2017. (2017). *World Economic Forum*. Retrieved April 4, 2020 from https://www.weforum.org/reports/the-global-risks-report-2017.

Human Rights Watch. (n.d.). Killer Robots. Retrieved April 1, 2020 from https://www.hrw.org/topic/arms/killer-robots.

IEAI. (n.d.). Responsible AI in Africa Network. Retrieved April 6, 2020 from https://ieai.mcts.tum.de/networks/.

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines. *arXiv preprint arXiv:1906.11668*.

Kakao Corporation. (2018). *Kakao Algorithm Ethics Charter (KAEC)*. Retrieved from https://www.kakaocorp.com/kakao/ai/algorithm?lang=en.

Kaloudi, N., & Li, J. (2020). The AI-Based Cyber Threat Landscape: A Survey. *ACM Computing Surveys (CSUR), 53*(1), 1-34.

Kirkpatrick, K. (2017). It's not the algorithm, it's the data. In: ACM New York, NY, USA.

Korea Agency for Infrastructure Technology Advancement. (n.d.). Collect Opinions on Self-driving Ethics Guidelines. Retrieved April 3, 2020 from https://www.kaia.re.kr/portal/contents.do?menuNo=200974 (Korean).

Korea Artificial Intelligence Ethics Association. (2019). *Charter of Artificial Intelligence Ethics (CAIE)*. Retrieved from https://kaiea.org/aicharter (Korean).

Korea Communications Commission, & Korea Information Society Development Institute. (2019). *Principles for User-Oriented Intelligence Society (PUOES)*. Retrieved from https://kcc.go.kr/user.do?boardId=1113&page=A05030000&dc=K00000200 &boardSeq=47874&mode=view (Korean).

Korea Law Information Center. (n.d.). *Korea Law Information Center*. Retrieved from http://www.law.go.kr/ (Korean).

Korean Internet Ethics Association, & National Information Society Agency. (2018). *Intelligent Government Ethics Guideline for Utilizing Artificial Intelligence (IGEG)*. Retrieved from http://www.alio.go.kr/popSusiViewB1040.do (Korean).

Kubovic, O., Kosinar, P., & Janosik, J. (2018). Can artificial intelligence power future malware? *ESET White Paper*, 1-15.

Kumagai, J. (2007). A robotic sentry for Korea's demilitarized zone. *IEEE Spectrum, 44*(3), 16-17.

Kumar, S. (2019). Advantages and Disadvantages of Artificial Intelligence. *Medium*. Retrieved April 6, 2020 from https://towardsdatascience.com/advantages-and-disadvantages-of-artificial-intelligence-182a5ef6588c.

Livingston, S., & Risse, M. (2019). The future impact of artificial intelligence on humans and human rights. *Ethics & International Affairs, 33*(2), 141-158.

Ministry of Commerce Industry and Energy. (2007). *Draft of the Robot Ethics Charter (DREC)*. Retrieved from https://cafe.naver.com/roboethics/8 (Korean).

Ministry of Land Infrastructure and Transport, & Korea Agency for Infrastructure Technology Advancement. (2019). *Ethical Guidelines for Self-driving Cars (EGSC)*. Retrieved from http://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?id=95083245 (Korean).

Ministry of Science and ICT, & National Information Society Agency. (2018). *Ethical Guidelines for Intelligence Information Society (EGIIS)*. Retrieved from https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=66361&bcIdx=20238&parentSeq=20238 (Korean).

Ohio State University. (2020). AI estimates unexploded bombs from Vietnam War: Machine learning detects bomb craters in Cambodia. *ScienceDaily*. Retrieved April 7, 2020 from www.sciencedaily.com/releases/2020/03/200324090005.htm.

Olson, P. (2018). This AI Just Beat Human Doctors On A Clinical Exam. *Forbes*. Retrieved April 9, 2020 from https://www.forbes.com/sites/parmyolson/2018/06/28/ai-doctors-exam-babylon-health/#4b31226012c0

PwC. (n.d.). How will AI change the future of work? Retrieved March 19, 2020 from https://www.pwc.com/us/en/services/consulting/analytics/artificial-intelligence/future-of-work.html.

R. AI Africa. (n.d.). KNUST AI NETWORK. Retrieved April 9, 2020 from https://raiafrica.org/.

Rohringer, T. J., Budhkar, A., & Rudzicz, F. (2019). Privacy versus artificial intelligence in medicine. *University of Toronto Medical Journal*, *96*(1).

Sedenberg, E., & Chuang, J. (2017). Smile for the camera: privacy and policy implications of emotion AI. *arXiv preprint arXiv:1709.00396*.

Surber, R. (2018). Artificial Intelligence: Autonomous Technology (AT), Lethal Autonomous Weapons Systems (LAWS) and Peace Time Threats. *ICT4Peace Foundation and the Zurich Hub for Ethics and Technology (ZHET) p*, *1*, 21.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, *59*(2), 204-217.

Wang, L. (2019). *The Subjective Value of Artistic Creation in the Age of Artificial Intelligence.* Paper presented at the 5th International Conference on Arts, Design and Contemporary Education (ICADCE 2019).

Zuiderveen Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making.