

A Study on Ethical Awareness Changes and Education in Artificial Intelligence Society

Jungin Kwon 

Department of Kyedang General Education, Sangmyung University, 20, Hongjimun 2-gil, Jongno-gu, Seoul 03016, Republic of Korea

Corresponding Author Email: jikwon@smu.ac.kr

<https://doi.org/10.18280/ria.370212>

Received: 14 February 2023

Accepted: 12 March 2023

Keywords:

artificial intelligence, artificial intelligence ethics, ethics education

ABSTRACT

In order to change our moral practice and contemplative consciousness during the change to the Artificial Intelligence society, Artificial Intelligence ethics education is necessary. Artificial Intelligence ethics education should aim to form moral human beings so that members of the Artificial Intelligence society can grow into moral subjects. Key elements of responsibility and safety, employment and discrimination, and tolerance and limitations were derived as core elements of Artificial Intelligence ethics education. Based on the derived core elements, the Artificial Intelligence ethics training course was constructed, and after the 14th week of learning, the change in learners' Artificial Intelligence ethics awareness was measured. As a result of the measurement, the improvement effect through Artificial Intelligence education was evident in responsibility and safety, tolerance and limit, but not in employment and differentiation. The purpose of this study is to present a direction for Artificial Intelligence ethics education by examining the educational values and limitations of Artificial Intelligence ethics education, and that Artificial Intelligence ethics education is necessary for members of the Artificial Intelligence society to grow into moral subjects.

1. INTRODUCTION

Recently, the whole world is keenly aware of how society will change due to the introduction of Artificial Intelligence technology in each field of society. Artificial Intelligence has the potential to our society, from the way we live and work to the way we communicate and interact with each other. It offers unprecedented opportunities for improving efficiency, productivity, and quality of life, and has already shown remarkable results in many fields. And there are discussions about how to properly regulate Artificial Intelligence that affects society without being excessive. This is because there is a possibility that excessive regulation on Artificial Intelligence will discourage the development of Artificial Intelligence and prevent society from obtaining benefits. However, the current development of Artificial Intelligence technology is in progress, and it is difficult to clearly prove the benefits or risks that society can gain from it, so the criteria for judging inappropriateness are very ambiguous [1].

For these reasons, current regulatory settings for Artificial Intelligence are emerging to create ethical guidelines for Artificial Intelligence worldwide. However, the premise of self-regulation, such as ethical guidelines, is the voluntary practice of the regulation. If voluntary practice does not take place, the very existence of regulation is meaningless. Therefore, discussions of enforceable regulation, the law on Artificial Intelligence, are emerging.

Bietti [2] argues that although many companies today talk about 'ethics' with Artificial Intelligence ethical principles, guidelines, and guidelines, when ethics exerts practical power in business operations is difficult to find. In short, Bietti [2] is raising a problem with the tendency to simplify and formalize

'ethics' as a rite of passage, where the term 'ethics' is unilaterally consumed for the reputation of corporations and governments. Some possible advantages of companies talking about ethics with AI principles, guidelines, and standards are that it could increase awareness and encourage discussions on ethical considerations in AI development and deployment. This could help companies make more informed decisions and avoid potential harm caused by AI systems. However, the shortcomings of such talk are that it may be only a public relations strategy to enhance the reputation of corporations and governments, rather than a genuine commitment to ethical AI practices. Furthermore, there may be a lack of clear implementation and enforcement mechanisms for these ethical principles, guidelines, and standards. Therefore, there is a need to improve the transparency and accountability of ethical AI practices, and to ensure that ethical considerations are embedded into the design, development, and deployment of AI systems.

This phenomenon is also affecting the education system. Ethical consciousness and necessity are recognized, but it is difficult to quantify actual results, so there is a tendency to treat ethics education as a part of education that simplifies and formalizes. However, we have already experienced in the past that society can become confused and perish if ethics are not equipped in using software as a tool [3].

Self-regulation of ethics or compulsory laws can be decided in a short time and set forth as rules, but our perception is different from rules or laws made in a short time. In order to change our moral practice and contemplative consciousness during the change to the Artificial Intelligence society, Artificial Intelligence ethics education is necessary [4, 5]. Artificial Intelligence ethics education should go beyond

simply learning technical prescriptions to cure the problems Artificial Intelligence brings to society and should aim at forming moral human beings so that members of the Artificial Intelligence society can grow into moral subjects. Under this direction, Artificial Intelligence ethics education should contain the process of making human Artificial Intelligence habituate moral practice and reflection based on understanding of ethics.

The purpose of this study is to present a direction for Artificial Intelligence ethics education by examining the educational values and limitations of Artificial Intelligence ethics education, and that Artificial Intelligence ethics education is necessary for members of the Artificial Intelligence society to grow into moral subjects. Through this, artificial intelligence can contribute to the establishment of a curriculum in a rapidly changing society.

AI ethics education should pursue the ability to use artificial intelligence properly and cultivate human morality. In addition, education should be provided for artificial intelligence and human beings to live in harmony. To this end, the goal is to nurture members of the artificial intelligence society so that they can think about purpose, data, and application areas, and grow into moral subjects.

2. THEORETICAL BACKGROUND

Artificial Intelligence ethics education is currently opened as various types of subjects in many countries around the world. In the case of Harvard University's embedded ethics course, it is characteristic that Artificial Intelligence ethics is integrated into the content of computer science, which is a major subject. The combined sub-fields include privacy issues in the content and design of large-scale distributed systems, human-computer interaction and system design for visually impaired users, discussion of machine learning and cases of inadvertent discrimination, computer networks and Facebook fake news issues. Also, Artificial Intelligence & Human Rights: Opportunities & Risks, presented by the Berkman Klein Center at the University of Harvard in 2018, dealt with issues of Artificial Intelligence and human rights in crime, financial system, health, education, etc. all [6].

Raso et al. [6] introduced an educational process to learn the bias of data and algorithms through An Ethics of Artificial Intelligence Curriculum for Middle School Students.

Also, the issues of Artificial Intelligence ethics emerging in Artificial Intelligence with autonomous learning functions such as deep learning. In 2017, the Future of Life Institute published the Asilomar Artificial Intelligence Principles. Its contents include safety, disability transparency, judicial transparency, accountability, alignment of values, human values, privacy, freedom and privacy, common interest, common prosperity, human control, non-destruction, and arms race as ethics and values [6, 7]. In the case of the European Union (EU), in 2018, it presented a trustworthy Artificial Intelligence ethics guideline consisting of contents such as legal, ethical, and robust. In April of the following year, 2019, Artificial Intelligence HLEG revised and announced the guidelines. As shown in Table 1 seven key requirements are presented to create a trust environment for Artificial Intelligence [8, 9]. The core requirements consist of human authority and oversight, technological robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, society and social well-being, and

accountability [10, 11]. The guideline presents a list of assessments that provide guidance on the actual implementation of each requirement [9, 12].

Table 1. Artificial intelligence system core requirements

Core Requirement	Details
Human agency and oversight	Artificial Intelligence systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches
Technical robustness and safety	Artificial Intelligence systems need to be resilient and secure. They need to be safe, ensuring a fall-back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.
Privacy and data governance	besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimised access to data.
Transparency	the data, system and Artificial Intelligence business models should be transparent. Traceability mechanisms can help achieving this. Moreover, Artificial Intelligence systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an Artificial Intelligence system and must be informed of the system's capabilities and limitations.
Diversity, non-discrimination and fairness	Unfair bias must be avoided, as it could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, Artificial Intelligence systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.
Societal and environmental well-being	Artificial Intelligence systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.
Accountability	Mechanisms should be put in place to ensure responsibility and accountability for Artificial Intelligence systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate an accessible redress should be ensured.

3. RESEARCH METHODS AND PROCEDURES

3.1 Research subjects and procedures

The subjects of this study were 104 students in Group A who did not choose Artificial Intelligence ethics courses and

112 students in Group B who took Artificial Intelligence ethics courses.

Students who will participate in the survey to measure Artificial Intelligence ethics were divided into two groups. The basic personal information of students is shown in the following Table 2.

Table 2. Research subjects

	Group A	Group B	Total
Male	48	60	108
Female	56	52	108
Total	104	112	216

N = 216

Table 3. Independent samples test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.998	.318	-.151	432	.880	-.049	.3267	-.691	.592
Equal variances not assumed			-.151	431.3	.8800	-.049	.3266	-.691	.592

In order to prove that groups A and B are the same group, a homogeneity test was conducted between the two groups prior to training. The tools used for the homogeneity test investigated Artificial Intelligence ethics based on the questionnaire items of Kim (2021) [13] Artificial Intelligence ethics test tool development research. The results of the homogeneity test of A and B groups (Independent samples test results: t-value: -.880, Sig (2-tailed): 0.880) are shown in Table 3. Thus, it was proved that the two groups are the same group.

Afterwards, 112 members of Group B conducted the Artificial Intelligence society ethics education course for 14 weeks in the first semester of 2022, and the detailed curriculum is shown in Table 4 [14, 15].

Group B provided a curriculum to recognize the importance of ethics in the Artificial Intelligence society based on the concept of Artificial Intelligence technology in connection with the subject of each unit of the ethics curriculum in Artificial Intelligence society.

In the curriculum, contents such as fairness of data, responsibility of information recipients and information providers, and prohibition of discrimination are emphasized.

Responsibility and safety education consists of discussion activities and reviews on the problems of self-driving cars and accident liability after activities on the Moral Machine site to understand the importance of responsibility and safety of Artificial Intelligence such as driverless cars.

The training on employment and discrimination was since 39% of people mistakenly identified as criminals by Amazon Rekognition, an Artificial Intelligence based facial recognition technology, appeared to be of color. In this regard, discussion activities and reviews were conducted on the bias of data and the dangers of bias in algorithms on race.

Education on permits and limits included discussion activities and reviews on the requirements for creation of copyright for digital content creations produced by Artificial Intelligence.

The study includes 104 students in Group A who did not take AI ethics courses and 112 students in Group B who did take AI ethics courses. The groups were tested for homogeneity before the course to ensure they were the same group. Group B took a 14-week AI ethics course that covered topics such as fairness of data, responsibility of information recipients and providers, and prohibition of discrimination. Discussions and reviews were conducted on topics such as responsibility and safety education, employment and discrimination, and permits and limits.

3.2 Research result

The results of the survey conducted on the two groups after the 14th week of ethics education are shown in the following Table 5.

Table 4. Artificial intelligence social ethics curriculum

Weeks	Subsection	Content of education
1 st	Anthropocentricity service	Changes in Artificial Intelligence society
2 nd	Stability and reliability of Artificial Intelligence technology	ICM-centered social change
3 rd	Respect for human basic values	Necessity of Artificial Intelligence Ethics
4 th	Legislation of Artificial Intelligence	Necessity of Artificial Intelligence legal system
5 th	Responsibility for Artificial Intelligence decision-making	Importance of personal information and ethical governance
6 th	Stability of Artificial Intelligence technology	Artificial Intelligence harmful content
7 th	Recognition of copyright	Artificial Intelligence copyright concept
8 th	Reliability of Artificial Intelligence technology	Case using Deep Fake
9 th	Differentiation in Data Access	Problems of Data Bias
10 th	Category for medical practice	Problems of Algorithms
11 th	Acceptable level of business ethics	Artificial Intelligence Ethics Policy
12 th	Differentiation of Artificial Intelligence decision making	Artificial Intelligence Ethics Guidelines
13 th	Artificial Intelligence User Responsibility	Responsibilities and Obligations of Developers and Acceptors
14 th	Necessity of Artificial Intelligence ethics education	Necessity of Artificial Intelligence ethics education

Table 5. Research results

	Subsection	Group A		Group B	
		Positive	Negative	Positive	Negative
Accountability and Stability	Responsibility of Artificial Intelligence developers	38 (36.5%)	66 (63.5%)	76 (67.9%)	36 (32.1%)
	Responsibilities of Artificial Intelligence users	42 (40.4%)	62 (59.6%)	61 (54.5%)	51 (45.5%)
	Responsibility of Artificial Intelligence decision-making	49 (47.1%)	55 (52.9%)	66 (58.9%)	46 (41.1%)
	Safety of Artificial Intelligence technology	51 (49.0%)	53 (51.0%)	56 (50.0%)	56 (50.0%)
	Reliability of AI technology	48 (46.2%)	56 (53.8%)	84 (75.0%)	28 (25.0%)
	Total	43.8%	56.2%	61.3%	38.7%
Employment and Discrimination	Differentiation of Artificial Intelligence decision-making	53 (51.0%)	51 (49.0%)	75 (67.0%)	37 (33.0%)
	Respect for human basic values	55 (52.9%)	49 (47.1%)	56 (50.0%)	56 (50.0%)
	Problems with job substitution	48 (46.2%)	56 (53.8%)	51 (45.5%)	61 (54.5%)
	Anthropocentricity service	52 (50.0%)	52 (50.0%)	60 (53.6%)	52 (46.4%)
	Differentiation of data access	51 (49.0%)	53 (51.0%)	45 (40.2%)	67 (59.8%)
	Total	49.8%	50.2%	51.3%	48.8%
Permissions and Limitations	Artificial Intelligence Legal Goods	42 (40.4%)	62 (59.6%)	79 (70.5%)	33 (29.5%)
	Recognition of copyright	35 (33.7%)	69 (66.3%)	93 (83.0%)	19 (17.0%)
	Recognition of Artificial Intelligence rights	25 (24.0%)	79 (76.0%)	89 (79.5%)	23 (20.5%)
	Categories of medical practice	40 (38.5%)	64 (61.5%)	74 (66.1%)	38 (33.9%)
	Tolerance of business ethics	51 (49.0%)	53 (51.0%)	71 (63.4%)	41 (36.65%)
	Total	37.1%	62.9%	72.5%	27.5%

Group A = 104, Group B = 112

Based on previous studies, this study divided the areas that are necessary in AI ethics education into three areas: Responsibility and Stability, Employment and Discrimination, and Tolerance and Limitations. Responsibility and Stability areas were again divided into AI developer responsibility, AI user responsibility, AI decision-making responsibility, AI technology safety, and AI technology reliability. The areas of employment and discrimination were divided into differences in artificial intelligence decision-making, respect for basic human values, problems with job substitution, human-centered services, and differences in data access. The areas of permissibility and limits were divided into legalization of artificial intelligence, recognition of copyright, recognition of rights of artificial intelligence, store owner for medical practice, and allowable value of business ethics. Afterwards, the need for artificial intelligence ethics education was measured for group A who did not complete AI ethics education and group B who completed AI ethics education, and the results were as follows.

As a result of comparing the two groups, group B was able to obtain higher average scores than group A in the domains of responsibility and stability, employment and discrimination, and tolerance and limits.

In the area of responsibility and safety, the importance of the responsibility of Artificial Intelligence developers and the reliability of Artificial Intelligence technology was high. Regarding employment and differentiation, the average of

group B was high, but the discrimination with group A was not so great.

In permission and limit, the importance of recognition of copyright and recognition of Artificial Intelligence rights was high, and it was also the area where the difference with group A was the largest.

Therefore, it was found that artificial intelligence ethics education is effective in instilling ethical awareness in the areas of responsibility and stability, tolerance, and limitation.

In short, the study conducted a survey on two groups of students, one with and one without artificial intelligence ethics courses. Group B, who received the ethics education, scored higher in responsibility and stability, employment and discrimination, and tolerance and limits. They showed a higher understanding of the responsibility of AI developers, employment and discrimination issues, and the importance of recognition of copyright and AI rights.

4. CONCLUSION

Artificial Intelligence ethics education should pursue not only the ability to use Artificial Intelligence correctly, but also the moral cultivation of humans who develop and use the technology in a balanced way. Artificial Intelligence ethics education should be an education for humans and Artificial Intelligence to live in harmony, in other words, an education

that fosters the ability for humans to live as the subject of life based on the desirable use of Artificial Intelligence.

To this end, the aspects of what is the purpose of the system and rules and legal values of society, what needs to be learned to achieve the purpose, and how to apply or utilize what has been learned must be considered.

Artificial Intelligence ethics education should present the following direction. First, it is necessary to deeply reflect on the purpose of creating Artificial Intelligence under the premise that the development and use of Artificial Intelligence is for human happiness. Second, in order to achieve the above purpose, it is necessary to think about determining the characteristics and types of data to be learned by Artificial Intelligence. Third, it is necessary to consider another area of application in which Artificial Intelligence can be appropriately used in the area for the benefit of the community as well as the individual.

This Study suggest that the development and use of Artificial Intelligence should be guided by ethical principles and considerations, in addition to technical knowledge and skills. The study highlights the importance of educating individuals who develop and use Artificial Intelligence to be morally responsible and cultivate a balanced relationship between humans and technology. So that the study emphasizes that Artificial Intelligence ethics education should be designed to help students reflect on the purpose of Artificial Intelligence, determine the types of data that should be learned, and consider the areas where Artificial Intelligence can be appropriately used for the benefit of both individuals and communities. It is suggested that such education should go beyond technical prescriptions for problems and focus on forming moral individuals who can make ethical decisions and reflect on their actions. Overall, the study suggests that Artificial Intelligence ethics education should promote the development of moral subjects who are capable of using Artificial Intelligence in a way that benefits society while ensuring that individuals' rights and dignity are respected. The findings of this study have important implications for policymakers, educators, and researchers working on Artificial Intelligence and its ethical implications.

Artificial Intelligence ethics education should go beyond simply learning technical prescriptions to cure the problems Artificial Intelligence brings to society and should aim at forming moral human beings so that members of the Artificial Intelligence society can grow into moral subjects. Under this direction, Artificial Intelligence ethics education should include a process that enables students to make moral practice and reflection a habit based on understanding of ethics in Artificial Intelligence.

REFERENCES

- [1] Lee, C.K., Oh, B.D. (2016). The robot ethics of the autonomous vehicle and its legal implications. *Hongik Law Review*, 17(2): 5-8. <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART002118649>
- [2] Bietti, E. (2020). From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, TUP, pp. 210-219. <https://doi.org/10.1145/3351095.3372860>
- [3] Lee, B.J. (2006). Artificial intelligence and the problem of responsibility. *Journal of the Daedong Philosophical Association*, 37: 73-92. <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART001037279>.
- [4] Lim, S. (2017). Moral education in the age of artificial intelligence: From the perspective of consumer ethic. *Journal of Ethics: The Korean Association of Ethics*, 1(117): 89-116. <https://doi.org/10.15801/je.1.117.201712.89>
- [5] Lee, C.H. (2020). Direction of software education in practical arts for cultivating competencies in the AI Era. *Journal of Korean Practical Arts Education*, 26(2): 41-64.
- [6] Raso, F.A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., Kim, L. (2018). Artificial intelligence & human rights: opportunities & risks. *Berkman Klein Center Research Publication*, (6). <https://nrs.harvard.edu/urn-3:HUL.InstRepos:38021439>, accessed on Jan. 10, 2023.
- [7] Future of Life Institute (2017). *Asilomar ai principles*. <https://futureoflife.org/ai-principles>, accessed on Jan. 3, 2023.
- [8] European Parliamentary Research Service (2019). *EU guidelines on ethics in artificial intelligence: Context and implementation*. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI\(2019\)640163_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf).
- [9] HLEG, A. (2019). *Ethics Guidelines for trustworthy AI*. European Commission, high level expert group on AI, pp. 1-39. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- [10] European Commission (2020). *Report from The Commission to The European Parliament, The Council and The European Economic and Social Committee. Report on The Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics*, European Commission, pp. 1-17. <https://op.europa.eu/en/publication-detail/-/publication/4ce205b8-53d2-11ea-aece-01aa75ed71a1/language-en>.
- [11] European Commission (2020). *White paper on artificial intelligence - A European approach to excellence and trust*. European Commission, pp. 1-26.
- [12] HLEG, A. (2019). *Policy and investment recommendations for trustworthy AI*, European Commission, pp. 1-50.
- [13] Kim, G.S., Shin, Y.J. (2021). Study on the development of a test for artificial intelligence ethical awareness. *Journal of The Korean Association of Artificial Intelligence Education*, 2(1): 1-19. <https://doi.org/10.52618/AIED.2021.2.1.1>
- [14] Ali, S., Payne, B.H., Williams, R., Park, H.W., Breazeal, C. (2019). Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education. In *International Workshop on Education in Artificial Intelligence K-12 (eduai'19)*, 2, pp. 1-4.
- [15] Williams, R., Ali, S., Devasia, N., DiPaola, D., Hong, J., Kaputsos, S.P., Jordan, B., Breazeal, C. (2022). AI+ ethics curricula for middle school youth: Lessons learned from three project-based curricula. *International Journal of Artificial Intelligence in Education*, 1-59. <https://doi.org/10.1007/s40593-022-00298-y>