

Robotics & AI Ethics

2021 6(2)

<Index>

1. South Korea's Strategic Culture and China's National AI STRATEGY: A Neoclassical Realist View.
/ **Chungsik Yu**
2. Future WARFARE and Artificial Intelligence Applications in the ROK Military: Problems and Alternatives.
/ **Jongha Kim**
3. Analyzing the Effects of AI Education Program based on AI Tools.
/ **Aeri Lee**
4. Mediating Effect of the Involvement with Beauty Content in the Relationship Between the Characteristics of YouTube Beauty Content That Uses AI and the Intention to Purchase Cosmetics: Comparison Between Men and Women from 20 to 40 Years Old.
/ **Hyangmee Woo, Eunsook Kim**
5. Approaches to Forming ETHICAL AI as an Artificial Moral Agent: Suggesting Virtue Education Method Through Comparison of Top-Down and Bottom-Up Approaches.
/ **Hyunsoo Kim**

Publisher: J-INSTITUTE
ISSN: 2435-3345

Website: www.j-institute.jp/ai/
Editor: ai@j-institute.jp

Corresponding author
E-mail: charlesyu917@naver.com

dx.doi.org/10.22471/ai.2021.6.2.01

© 2021 J-INSTITUTE

South Korea's Strategic Culture and China's National AI STRATEGY: A Neoclassical Realist View

Chungsik Yu

Shanghai Jiaotong University, Shanghai, China

Abstract

Purpose: South Korea's foreign policy executives need to make security policy choices about China's AI as a security threat with normative nature. The purpose of this article is to identify the ideas of the South Korean security policy executives and to analyze the role of the policy elites' perception in foreign policymaking about China's artificial intelligence as a newly emerging security challenge. Together with this, it is approached to evaluate the strategic perception of Moon Jae-in's foreign policy executive in connection with China's artificial intelligence.

Method: The article examines the document and speech of the foreign policymakers of Korea and China as well as the media coverage of the two countries by utilizing Kitchen's neoclassical realism with a focus on the elites' perception as an intervening variable, international systemic pressure as an independent variable, and foreign policy outcome as a dependent variable. Chapter 2 reviews the strategic culture of a middle power. Chapter 3 finds China's AI strategy as a systemic threat. Chapter 4 explores the former Korean governments' strategic ideas for China. Chapter 5 finds Moon Jae-in government's idea of the China AI strategy.

Results: The article reveals the nature of China's AI, the role of Korea's strategic culture, and the capacity of policy elites to emerging security. First, China's AI has the nature of systemic pressure. Second, the foreign policy executives of Korea differently respond to China's systemic pressure. Third, the Moon Jae-in government's policy elites have autonomous power against the international systemic pressure and consider the economic interest as important.

Conclusion: This article concluded that the foreign policy elites' perception intervenes between the international systemic pressure and state foreign policy-making. The elites induce the discussion of AI from the view of middle power multilateralism rather than that of AI from the view of ethical and geopolitical context. The article also says that Korea's foreign policy elites can operationalize the ideas in considering the balance of interest with China.

[Keywords] AI Strategy, Emerging Security Challenge, Systemic Pressure, Strategic Culture, Middle Power

1. Introduction

China's national artificial intelligence strategy has been perceived as a newly emerging security threat to surrounding countries as well as great powers. The US foreign policy leaders see China's AI as a game-changer that makes it possible for China to overcome the inferiority of conventional weapons and nuclear capacity and to finally surpass the hegemon. Additionally, Japan, Taiwan, and Indonesia eye with care the China's rise in AI that serves as a geopolitical destabilizer to trigger a cyber arms race in the East Asian region. However, South Korea's government is relatively tranquil, compared with the states. The government endeavors to catch up with China in the arena of business-oriented and consumer-centered AI technology, but is reluctant to acknowledge China's AI as an emerging security

threat. We need to seek an answer about the reason why the Korean foreign policy elites do passively issue a question on China's AI as a security threat.

The central proposition of this article is that the strategic culture of South Korea's foreign policy elites' perception induces the discussion of AI from the view of middle power multilateralism rather than that of AI from the view of ethical and geopolitical aspects. The proposition is supported by the role of strategic culture in foreign policymaking that the stimuli of an international system do not directly affect the foreign policy-making of the state. Rather, the idea of policy elites comes between international systemic pressure and state security policy. Then, the policy elites interpret the system stimuli for themselves and choose a policy among multiple political options under limited resources to be domestically mobilized. Accordingly, the foreign policy of a state is shaped by foreign policy elites' idea that selects a foreign policy option in the strategic cultural environment.

This article aims to identify the ideas of South Korea's security policy executives to China's AI and to analyze the role of the policy elites' perception in foreign policymaking by taking Kitchen's neoclassical realism, in which strategic, operational ideas serve as the factor to integrate materiality and immateriality, and system and unit in international politics. This article is organized as follows. I begin with a review of the literature on emerging security challenges, systemic threats, and domestic constraints, and policy elites' perception, and middle power. Through the contents analysis of political party and government documents, top leaders' speech, and media coverage, I argue that China's Korean peninsula policies are differently accepted by South Korea's foreign policies executives, so is China's AI. This is because the elites can operationalize the ideas in considering the balance of interest with China.

2. Systemic Threats and the Strategic Culture of Middle Power

2.1. AI as an emerging security challenge and balance of power

A pair of the conventional lens of security does not capture the nature of artificial intelligence with a respect to a newly emerging security threat. Put simply, one is a state-centric approach arguing that threats come from an interstate military use of weapons. The other is an inter-subjective approach contending that threats arise from humans' consciously and reciprocally conceived object. By criticizing the conceptual narrowness of traditional security, Herd, Puhl & Costigan suggest emerging security threat as an inclusive concept, which denotes an event that puts at risk the material or the immaterial basis on which individual, society, state, and the earth depend[1]. Johnson also focuses on numerous challenges on a global scale[2]. However, they do not fully touch upon the geopolitical and moral dimension of emerging security threats, albeit their inclusive definition.

AI as a newly emerging threat in the international system is a major source to alter geopolitical power equations[3]. Many theorists say that global hegemon, aspiring hegemons, and middle powers jump into the arms race of the cyberspace weapons based on AI-enabled weapons, which gives rise to the disequilibrium of the balance of power in the international system[4]. G20 countries, involving Canada, China, France, Germany, India, Japan, Russia, South Korea, the UK, Israel, and the US massively invest expenditure for AI technology development. Among those countries, eight are in the Asia-Pacific region. Moreover, China aspires to be a global leader in 2030. AI-enabled weapons systems as an accelerator pedal of US-China technological hegemony competition exacerbates security dilemma, leading to an imbalance of power in the East Asian region.

2.2. Systemic pressure and domestic constraints: neorealism and neoclassical realism

Given that AI is a structural factor to accelerate the arms race among states, we need to look into the logic that threats stem from the international system itself. According to neorealism, international systemic pressure constrains state actions such as state-leveled security policymaking. For the structuralist, state behavior is a dependent variable because the international anarchy, with no central governing body, compels states to survive with an exclusive reliance on self-help and to be functionally undifferentiated. States differ from each other only by the difference in material capabilities that

contribute to their safety. The inter-state distribution of material capabilities is an independent variable that brings about the change of the international system such as war and peace, alliance, and balance of power. The inter-state competition arising out of AI-driven material capabilities is also, for neorealist, regarded as the systemic pressure constraining state action[5].

Neoclassical realists criticize that the power distribution perspective does not grab full attention to state action that affects foreign policy outcome. According to the neoclassical realist, the neorealist wrongly asserts that international distribution of material capabilities does unilaterally affect foreign policymaking. For the neoclassical realist, the condition of a state's relative power in the international system can affect its foreign policymaking, but the state unit-level factors can also affect the relative power distribution of a state. Thus, neoclassical realist argues that all the nation-states do not make an equal foreign policy response to the international system-level pressure because the factors affecting state behavior can be different from state to state. That is to say, intervening factors interpose themselves between international systemic pressure and policy outcome, and finally perceive the systemic threat from their point of view.

The intervening variables are internal factors arising from within the state, which affect the foreign policymaking process. The examples are domestic constraints and state agency ability. The former concerns policy issues and state-society relations, and the latter is policymakers' ideas and perceptions[6]. The ideas involve economic interest groups' orientation and state policymakers' perception of foreign policy that is formed for a long time. Such domestic constraints are not only affected by the international system but also affect the international system. Accordingly, neoclassical realists need to reveal how domestic constraints and state policy actors influence the international structure.

2.3. Policy elites' perception as strategic culture

Taking a hint from the constructivists' idea, a group of neoclassical realists refutes that international systemic pressure has a unilateral influence upon foreign policy outcome. Instead, neoclassical realists propose the concept of a strategic idea as an intervening variable between international systemic pressure and the foreign policy decision-making process[7]. Theorists of strategic culture put special attention to policy idea-set, habitual patterns, and historical motivation that influences security policy choice[8]. In strategic culture, the key is state foreign policy elites' perception and domestic economic forces even within the predefined geopolitical tension. Foulon rightly points out that strategic idea serves as a bridge between material-focused and idea-emphasized explanations in international security[9].

However, Foulon's discussion on the state-level policymakers' perception and domestic commercial interest mainly focuses on great powers such as De Gaulle's foreign policy in the 1950s and the US' recent rebalancing to East Asia. Namely, neoclassical realists concentrate on debating with Waltzian neorealism with a focus on great powers and insufficiently analyzed the policymakers' perception and domestic interest group's pressure for non-great powers. Currently, the East Asian region shows a rising instability of the geopolitical power equation since China adopted AI as the national strategy for a military build-up. Neoclassical realism needs to extend its theoretical application to the perceptual layer at the state policymakers and domestic constraints in the non-great powers of East Asia. Particularly, the analysis of foreign policy elites' perception in the context provides an opportunity to extend the theoretical appropriateness of neoclassical realism, providing that there has been a tradition of the strong state and weak society in the region.

2.4. Middle power and strategic culture in the international system

The concept of middle power shows a logical similarity to that of strategic culture in that idea plays a major role in forming state action. As Massie pinpoints, a myriad of policy elites' perception turns into a policy group identity in a long decision-making process. According to her, the identity consists of the internal one conceived by the citizens of the country as well as the external one acknowledged by other countries' people[10]. Similarly, for a middle power to be recognized as a middle body in the international system, the middle power, for oneself, need to identify as the middle one as well as other states perceive the middle power as a middle body in the international system. A middle-sized state

can be similar to middle power as regards the material ability. It is not a middle power, however, if the middle-sized state does not construe itself as middle. This point reminds us of the constructivist Wendt saying, 'Anarchy is what state makes of it.' In the same token, we can talk as follows: 'middle power is what the state makes of it.'

There are three views on what makes the state be seen as a middle power in the international system; that is to say, positional, stylistic, and self-identical approaches. First, the positional approach views material hierarchy as critical. How much material capability a state has determines itself as a middle-ranking state in the international hierarchy. Second, the stylistic approach regards the state's diplomatic behavior as essential. A typical and repeated way of managing the relations among countries, plus material ability, renders the state look like middle power. Third, the self-identical approach considers as important recognizing itself to be a member of a particular group. States in the third category are substantially placed as a middle-position in the hierarchy and at the same time perceive and distinguish themselves from other states.

Recently, there has been a keen interest in the ideas of middle powers' foreign policy executives with a respect to AI threats. The first trend of the study is that the development of AI technology exacerbates the risk of inadvertent war and the security dilemmas between powers as in the examples of the U.S. and China[11]. The second trend of the study concerns the relationship between middle powers and policy elites' perception. The study mainly discusses the effect of identity norms on the alliance between middle powers, national security strategy, the difference among middle powers in policy dimensions, and finally a decision-making and path dependency[12].

Despite the contribution to the middle powers and strategic culture, the extant studies neglect how the policy elites' perception of middle powers tackles the threats from an aspiring hegemon's artificial intelligence in regional politics. Further studies need to direct any attention to the research gap in the concrete case-setting of middle powers in the East Asian region. Drawing on the research gap, the article poses the following questions; the first is how do we understand China's national AI strategy in the international system? The second is what are the strategic ideas of the foreign policy executives of the middle power Korea? The third is what does Moon Jae-in government's diplomatic response to China's AI mean?

3. China's AI Technology as a Systemic Threat for Neighboring Countries

China's national AI strategy has a nationalistic aim to be a regional hegemon competing with the US and finally replacing the position of her counterpart in East Asia. China lagged behind the US in conventional and nuclear weapons. Since the inception of the so-called 'China's dream' policy, the Chinese leadership sought a novel strategy to outpace the world hegemon US. The leadership set out the national AI strategy as 'leapfrog strategy', denoting that actors falling behind competitors devise a drastic innovation that makes it possible for the latecomer to overtake its rivals and finally to be a strong socialist state[13]. The tradition of party-state, where a paramount leader or collective leadership makes critical policymaking, also fits into the national AI strategy of China[14]. Such tradition is found in the party and state leaders' official talks.

The paramount leader Xi Jinping in 2014 first proposed the military use of AI-enabled weapons as independent foreign policy focusing on the great rejuvenation of the Chinese nation.' In the report named 'Made in China 2015', the premier Li Keqiang emphasized the pivotal role of AI that plays in leading all the manufacturing. Furthermore, the report emphasized the 'strong' manufacturing state, the ambitious aim of which is to ultimately enter a global-leveled manufacturing state. I analyze the procedural aims as three; first, China tries to optimize the AI development environment to adjust itself to compete with other powers by 2020, second, China tries to be a world leader in some AI applications in 2025, and third, China tries to establish a world innovation center of AI that renders China be the technological hegemon in 2030. Thereafter, China in 2017 proclaimed 'New Generation Artificial Intelligence Development Plan(新一代人工智能发展计划, AIDP). The aim of AIDP, for China, is to become the world leader in 2030 and apply AI technologies to various areas.

There is a need to comprehend China's grand AI strategy from the angle of US-China technological hegemony competition. According to the power transition perspective, China aims to be an aspiring hegemon constructing the new bipolar international system, finally replacing the position of US superpower with that of China. In this context, the actual deployment of the technology based on China's national AI strategy is understood as the intention of China's leadership that modifies the geopolitical power equilibrium. Naturally, the deployment of AI technology brings about a systemic stimulus that deepens mutual distrust in the East Asian region, serving as a destabilizing factor in the geopolitical power equation.

The application of China's AI technology to military affairs is also grasped as an emerging security threat with moral and ethical vulnerability. Emerging security threat putting in danger the material or immaterial ground of multiple actors' life brings about moral and ethical controversies on a global scale. At present, there is no clearly defined norm about how the state should behave while using AI military weapons. If the actor of regime attribute is considered in international politics, China has been perceived as an illiberal and authoritarian state that is neglectful of the democratic controls of AI military weapons. Together with the nature of the international systemic threat, China's AI with the nature of emerging security threat can play as mounting tensions in international politics.

Through the national AI strategy, China aims at becoming AI regional hegemon in East Asia. The strategy is a variable to affect the regional stability and instability around the neighboring countries that are the main middle powers in the international system. Among them, South Korea is a country adjacent to China and depends upon the aspiring hegemon for the export of Korea's industrial goods. How has South Korea responded to the systemic stimuli?

4. Former South Korean Governments' Strategic Culture to China

Be based on the proposition, 'Bringing the state back in' made by neoclassical realists, this article sets systemic stimulus as an independent variable, and policy elites' perception and domestic constraint as intervening variables. The elites' perception is the strategic ideas of foreign policymaking executives, and the domestic constraints are the demands of Korean big business to the foreign policymakers.

I present the research result in advance that is gained through the analysis of the collected data. There exist international systemic stimuli that come from China's AI strategy in the context of US-China hegemony competition. The systemic pressure applies equally to all the four foreign policy executives of South Korea. However, the four foreign policy executives perceived the conflict between the US and China differently. Additionally, the four executives proposed strategic ideas that are distinct from each other. Around the systemic pressure, the common and different strategic ideas between the governments run as follows[15].

Table 1. Strategic ideas of foreign policy executives.

Variable			Foreign Policy Executive			
			Roh Moo-hyun	Lee Myung-bak	Park Geun-hye	Moon Jae-in
Independent variable	Systemic pressure	Degree	Low	High	High	High
		US policy to China	Engagement & containment	Engagement & containment	Containment	Containment
		Result	Cooperative competition	Rivalry between rebalancing to Asia & new type of great power relationship	Rivalry between rebalancing to Asia & new type of great power relationship	America's checking and blocking China
		Instability	Nuclear test of N. Korea	Nuclear test of N. Korea	THAAD	AI & pandemic
Intervening	Domestic constraints	Emphasis on	Emphasis on	Emphasis on	Emphasis on	

variable		China market	China market	China market	China market
	Elites' perception	National independence consciousness Breakaway from subordination to powers	Korea's stark alliance with the US Pro-US tendency	Conciliatory strategy Oscillating between China & US	Resistant nationalism Community of destiny in economy with China
Dependent variable	Foreign policy outcome to China	Comprehensive cooperative partnership	Rhetoric strategic partnership	Peninsula trust process	Korean peninsula driver theory
	Foreign policy outcome to East Asia	Korea as a balancer in Northeast Asia	Pragmatic diplomacy	Strategic ambiguity	Normative balancing

4.1. Systemic pressure from the US-China rivalry

The systemic pressure as an independent variable arises out of the competition over international power distribution in a neorealist sense. Here, the international distribution affecting Korea's foreign policy refers to the material capacity differently allocated between the US and China. Presumably, the confrontation to earn more relative gains between the two powers can be influential in Korea's foreign policy. The higher the degree of competition between the US and China is, the higher the degree of systemic pressure is. Conversely, the lower the degree of competition between the US and China is, the lower the degree of systemic pressure between the two powers is. The rivalry between the two powers is persistently high, although it may change more or less, depending on the foreign policy leaders of the two powers. For the South Korean foreign policy, the level of systemic pressure from the competition remained high, except for the period of Roh's government.

During the period of Roh's government, there is a low degree of competition between the US and China in East Asia. The US launched an engagement policy toward China, and China negatively responded to the US' unilateralism. There was an economic symbiosis between the two powers. The instability factor to increase the pressure was North Korea's nuclear test. By the way, the remainder of the three governments underwent the high systemic pressure from the competition between the two powers. During Lee's government, the pressure was heightened because there started a confrontation between the US' rebalancing to Asia' and China's new type of great power relationship in the regions. The other instability factor was North Korea's nuclear test[16]. During Park's government, the pressure was intensified as well. The instability factor of the government was THAAD(Terminal High Altitude Area Defense) stationed in South Korea. Finally, the systemic pressure during Moon's government is also high. The US containment policy toward China, like in the case of Park's government, continued. The Trump government started a hardline containment policy to China, America's checking and blocking China[17].

4.2. Domestic constraints: state-market relations

Domestic constraints refer to the relationship between the Korean government and business. All the foreign policy executives of the four governments laid the stress on the China market. The governments' emphasis results from the demand of the large and small businesses that massively export products to China. South Korea, a typical trading country, positively accepted the demand from business. The relationship between government and business as an intervening variable affected both international system and state action by coming between the international system and state foreign policy-making process.

4.3. Elites' threat perception on China

If we apply the system pressure variable to the four foreign policy executives, we find that there is a behavioral characteristic peculiar to the foreign policy executives. Korea's foreign policy executives

differently responded to China's AI as a security threat. Roh and Moon put more emphasis on a cooperative relationship with China. First, Roh had a national independence consciousness for peninsula nation-state between major powers and pursued a breakaway from subordinate relationship to the powers. Second, Moon, a supporter of resistant nationalism identity, perceived China as a pivotal economic partnership that was represented as a 'community of destiny' in economy with China. The two presidents acknowledged China as a substantially strategic partner in the Korean peninsula problem as well as in the economy.

In contrast, Lee and Park took unfriendly or inconsistent posture to China. First, Lee pursued pro-American policies and believed that South Korea should make a stark alliance with the US. The president strengthened the relationship between the US and South Korea, though he clearly articulated about making the strategic partnership with China that is considered just as rhetoric speech. Second, Park fluctuated between pro-China and anti-China. Initially, the top foreign policymaker adopted a conciliatory 'wait and see' attitude to China to draw China's support to solve North Korea's nuclear arms. However, the president, with the station of THAAD, took a drastically anti-China policy. The policymaker's turn was criticized as unprincipled oscillating between China and the US. The two presidents took an antagonistic or appeasing posture to China.

4.4. State action: foreign policy outcome

Consequently, the strategic idea of each foreign policy executive intervenes between the international system stimuli and foreign policymaking. The foreign policies are formed, relying on the four top-policy makers. Roh pursued a comprehensive cooperative partnership with China and perceived Korea as a regional balancer between the two powers in the security agenda. Conversely, Lee's perception of China is negative and took rhetoric strategic partnership based on pragmatic diplomacy. Park's regional foreign policy principle is strategic ambiguity. Her foreign policy to China was the trust-building process that means the Korea-US alliance and Korea-China strategic cooperation(聯美和中). Moon's foreign policy to China is Korean peninsula driver theory where South Korea plays a leading role in mediating the interest between the two powers. Besides, he argued the two Koreas should play a subjective role in regional international politics. The normative balancing of Moon as regional foreign policy is that South Korea pursues a more reconciliatory posture for the unification and mediates the regional interests based on Norm-based balanced diplomacy.

5. Moon Jae-in Government's Strategic Ideas to China's National AI Strategy

Why do not Moon's foreign policy executives raise the ethical riskiness or geopolitical impact that China's AI brings about? Kitchen's neoclassical realism gives a clue to the puzzle. According to Kitchen, there are three ideas; scientific, intentional, and operational. The ideas play as an intervening variable between structural pressure and foreign policy and can convert the pressure into concrete policy outcome[18]. First, the scientific idea concerns how the world works. The ideas set a range of choices where policymakers can opt for a preferred strategy. Second, intentional idea concerns what foreign policy should be selected. The intentional idea with normative nature cannot be preferred by policy elites if it damages national interest. Third, an operational idea is an idea that can be altered as scientific or intentional, according to the possibility of whether or not a goal is achieved.

For Kitchen, policy elites perceive threats through their operationalization of threats. According to intentional ideas, foreign policymakers should propose an ethical criticism to the use of China's AI because of the possible immoral results that AI-enabled killer robots and lethal autonomous weapons(LAWS) kill humans. However, Moon's foreign policy elites do not criticize China's AI. For Kitchen, this is because the ideas operationalized by foreign policy elites intervene between the systemic pressures and foreign policymaking. Namely, China's AI is operationalized as a scientific idea by the elites although the AI of China is the object of the intentional idea. The elites consider AI as the boundary of scientific ideas where the world works. AI as a scientific object is an object to analyze how it works, not to judge it as a standard of ethics. The policy elites are to operationalize an idea to be judged into an idea to be analyzed[19].

The strategic idea formed through the operationalizing process intervenes between the international systemic pressure from China's AI strategy and the foreign policymaking process, finally embodying the operationalized scientific idea into a foreign policy outcome. In June 2020, Moon's government, for the first time in the world, launched 'Global Partnership on Artificial Intelligence(GPAI), a multilateralism-oriented multi-stakeholder governance mechanism to manage the risk of AI[20]. I argue that the GPAI is a combined result between an operationalized scientific idea and the strategic ideas formulated as middle power identity.

6. Conclusion

This study reviews the strategic culture of middle power's policy elites with a respect to China's AI strategy as an emerging security challenge and reveals the role of the policy elites' ideas that interpose themselves as an intervening variable that affects both systemic pressure and policy outcome. The theoretical perspective is Kitchen's neoclassical realism, arguing that policy ideas connect the international system and state unit as well as material capability and foreign policy outcome. The Korean foreign policy executives respectively formulated different and idiosyncratic norms to the identical systemic pressure from the US-China hegemony competition and China threats.

By applying the role of the Korean policy elites' idea to China's AI strategy, this study extends Kitchen's neoclassical realism on middle powers. The study contributes to strategic culture studies. First, the findings of the study suggest that the conventional security threats such as human's identified security and military security threats are limited to encompassing the emerging security threats as a unit of analysis. Thus, the new security agenda such as AI and pandemic can be used as a new independent variable as an embodiment of international systemic pressure. Second, this study proves that the perception of state policymakers can be autonomous from the pressure of structure by showing that a plurality of South Korea's foreign security executives differently responds to the identical pressures of the international system. In this sense, the findings reveal that idea of policy elites is an intervening variable to cause the autonomy of state action.

Third, the strategic culture of policy elites used as an intervening variable illustrates the limitation of structural realism that depends upon the power distribution of the international system as the independent variable. By revealing that immaterial being can affect the material distribution, the findings say that the results of the policy as a dependent variable do not necessarily reflect material interests. Fourth, the findings show that the identity of policymakers can be used as a bridge between material capability and policy outcome as state action. This point presents that idea as an intervening variable that can affect both international system and state action and can be affected by the international system and state action. This also shows that the strategic idea of policymakers can be an underlying motif that shows the reason why one specific policy is adopted.

This study contributes to the judgment of foreign policy executives, who try to understand the role of the elites' perception in the foreign policymaking. By looking at the study findings, the top foreign policymakers can understand that the formation of strategic culture comes from policy elites' ideas. Together with it, normative ideas can play a critical role in the foreign policymaking of a middle power. However, this study has a limit in research methods. In data collection, the study only utilized government published documents and media storage. In a further study, there is a need for in-depth interviews of foreign policy executives. Additionally, there is a need for a comparative study between traditional and emerging middle powers with focus on the ideas of policy elites.

7. References

7.1. Journal articles

- [1] Shea J. How is NATO Dealing with Emerging Security Challenges? *Georgetown Journal of International Affairs*, 14(2), 193-201 (2013).

- [2] Johnson J. Artificial Intelligence & Future Warfare: Implications for International Security. *Defense & Security Analysis*, 35(2), 147-169 (2019).
- [3] Cho H & Park S. Human Security Task and Prospect on Covid-19 Pandemic. *International Journal of Terrorism & National Security*, 5(2), 48-55 (2020). [Article]
- [4] Li Y & Park G. AI Ethics and Privacy Right. *Robotics & AI Ethics*, 5(2), 27-33 (2020). [Article]
- [5] Choi W & Moon H & Shin J. A Study on the Overcoming of the Asymmetry of the Korea-U.S. Alliance: Focusing on the Trump Administration's Security, Economic and Military Alliance Policies. *International Journal of Military Affairs*, 4(2), 1-9 (2019). [Article]
- [6] Jeffrey WT & Lobell SE & Ripsman NM. Is Peaceful Change in World Politics Always Desirable? A Neoclassical Realist Perspective. *International Studies Review*, 20(2), 283-291 (2018).
- [7] Lantis JS. Strategic Cultures and Security Policies in the Asia Pacific. *Contemporary Security Policy*, 35(2), 166-186 (2014).
- [8] Nam J & Jo S. A Study on Canada's Response Systems for Terrorism and Emerging Trends. *International Journal of Terrorism & National Security*, 5(1), 13-24 (2020). [Article]
- [9] Foulon M. Neoclassical Realism: Challengers and Bridging Identities. *International Studies Review*, 17(4), 635-661 (2015).
- [10] Massie J. Making Sense of Canada's Irrational International Security Policy: A Tale of Three Strategic Cultures. *International Journal*, 64(3), 625-645 (2009).
- [11] Johnson JS. Artificial Intelligence: A Threat to Strategic Stability. *Strategic Studies Quarterly*, 14(1), 16-39 (2020).
- [12] Beeson M & Bloomfield A. The Trump Effect Down Under U.S. Allies, Australian Strategic Culture, and the Politics of Path Dependence. *Contemporary Security Policy*, 40(3), 335-361 (2019).
- [15] Joo Z. Roh Moo-hyun and Lee Myung-bak Governments' Policy toward China. *Sinology*, 30, 431-468 (2008).
- [16] Byun C. Evaluation on the Lee Myung-bak's Pragmatism and the Foreign Policy toward China. *Unification Strategy*, 13(1), 163-190 (2013).
- [17] Park Y. US-China Rivalry and the Moon Jae-in Government's Strategy. *Journal of North Asian Studies*, 25(1), 5-27 (2020).
- [18] Kitchen N. Systemic Pressures and Domestic Ideas: A Neoclassical Realist Model of Grand Strategy Formation. *Review of International Studies*, 36(1), 117-143 (2010).
- [19] Na H & Choi G. China's Revolution in Military Affairs based on Artificial Intelligence: Focusing on Military Intelligentization. *Korean Military Studies Review*, 76(3), 91-118 (2020).

7.2. Additional references

- [13] 熊辉. 人民论坛. 第31期 (2017).
- [14] 蔡自兴. 中国人工智能40年 科技导报 (2016).
- [20] Kim M. The World's First Global Partnership on Artificial Intelligence(GPAI) Officially Established. Pintech Economy Newspaper (2020).

8. Authors Contribution

8.1. Authors contribution

	Initial name	Contribution
Author	CY	<ul style="list-style-type: none"> -Set of concepts <input checked="" type="checkbox"/> -Design <input checked="" type="checkbox"/> -Getting results <input checked="" type="checkbox"/> -Analysis <input checked="" type="checkbox"/> -Make a significant contribution to collection <input checked="" type="checkbox"/> -Final approval of the paper <input checked="" type="checkbox"/> -Corresponding <input checked="" type="checkbox"/> -Play a decisive role in modification <input checked="" type="checkbox"/> -Significant contributions to concepts, designs, practices, analysis and interpretation of data <input checked="" type="checkbox"/> -Participants in Drafting and Revising Papers <input checked="" type="checkbox"/> -Someone who can explain all aspects of the paper <input checked="" type="checkbox"/>

Publisher: J-INSTITUTE
ISSN: 2435-3345

Website: www.j-institute.jp/ai/
Editor: ai@j-institute.jp

Corresponding author
E-mail: jong-ha44@hanmail.net

dx.doi.org/10.22471/ai.2021.6.2.11

© 2021 J-INSTITUTE

Future WARFARE and Artificial Intelligence Applications in the ROK Military: Problems and Alternatives

Jongha Kim

Hannam University, Daejeon, Republic of Korea

Abstract

Purpose: This article explores the most likely impacts of Artificial Intelligence(AI) on the character of future warfare, focuses on how AI will change just about the way we fight in the battlefield, analyzes the current status and problems of AI applications in the ROK military, and then suggests several policy alternatives with regard to future development of the ROK military's war-fighting capabilities with AI and autonomous systems (i.e., robots).

Method: This article uses a literature review method that collects and analyzes data through existing research results or literatures. In order to understand the research trends of AI-driven warfare, professional military papers related to future operational concepts and doctrines are analyzed. In addition, data published by the Army, Navy, and Air Force Headquarters is referenced for the current state and problems of AI application of the ROK military. In particular, the latest data is mainly reflected in order not to miss important issues related to the rapidly changing military application of AI.

Results: AI-driven future warfare will have three broad trends, that is, the acceleration of multi-domain battle, the generalization of cognitive-centered operations, and the Acceleration of Human-machine Fusion and increased ambiguity of blurring distinction between combatants and non-combatants. AI applications in the ROK military, however, is somewhat behind advanced countries, and currently stays at the conceptual level and exists mostly on paper. It still does not act to convert AI technology for military use, and not adapt it for operational concepts and acquisition process of weapons systems. For this reason, the ROK military needs to actively pursue various AI technology development strategies that can be used in conducting complex war-fighting within the full spectrum of conflicts.

Conclusion: The ROK military needs to discard ineffective legacy weapons systems and prioritize investment in advanced technologies including AI, and streamline its cumbersome weapons acquisition and sustainment process and procedure as a means to win the AI race with other countries. With AI, the ROK military will be able to create more agile, flexible and lethal war-fighting capabilities in strategic, operational, and tactical dimensions.

[Keywords] Artificial Intelligence(AI), Weapons Systems, Future Warfare, Operations, ROK Military

1. Introduction

In the past two decades, most countries across the world have been increasingly convinced of the importance of Artificial Intelligence(AI) in their societal development. That is especially true with regard to the impact of AI on rapid changes in industrial structure and social institutions. AI technologies are inherently dual-use, hence are being integrated into an increasing number of commercial and military applications. Due to the significant technological attributes, many countries, particularly US, China, and Russia, are currently investing heavily in the development of AI and autonomous systems(AS) at the national strategy level[1].

In this process, since the weaponization of AI is inevitable, there is an urgent need to build an appropriate global governance system[2] - international norms on AI - to ban the use or development of lethal autonomous weapons systems powered by AI software, such as 'killer robots.' Pfaff indicates

that “reluctance to develop AI applications for military purposes is not going to go away as the development, acquisition, and employment of these systems challenge the traditional norms associated with not just war-fighting but morality in general”[3].

Indeed there is a possibility that lethal autonomous weapons systems powered by AI software are likely to fall into a security dilemma, and thus will encourage AI arms race. However, it is slightly different from that of the nuclear arms race. First, unlike nuclear weapons, AI technology can be used not only in private sector but also in military areas, meaning the wider cast of actors exists, hence it is really harder to control. Second, technology transfer and investment are actively carried out through cross-border exchanges between private sectors, and international cooperation is becoming common. And, third, most AI-enabled weapons systems are still relatively immature and not yet widely deployed. It will take a considerable amount of time to become deeply embedded in military systems, given that its research and development(R&D) speed has been slow in comparison with the technological progress in the commercial sector. For these reasons, there is no need to stress too much the negative aspects of AI technologies by adding the term so-called ‘arms race.’

At any rate, the accelerated military applications of AI can be explained in straightforward terms. There is the consistent pressure to improve or develop weapons systems to match the level likely to be attained by one’s potential adversaries. That is, the obligation to keep weapons systems ahead of adversaries leads to a consistent demand for development of AI-driven new weapons systems. Indeed, despite ethical considerations of AI, both governments and defense industries, as main purchasers and producers of weapons systems, have a common interest in generating AI-driven innovative weapons systems which will keep their militaries ahead of the most proximate peer competitor. Given this common objective, and given that no one benefits from making low-quality weapons systems, one might expect that investment in AI technology is readily understood.

The current use of AI in the military is seen mainly in terms of ‘war-fighting capabilities,’ or ‘force multiplier’ that permits fielding of a relatively smaller military force while maintaining a decisive military advantage in both traditional and non-traditional warfare. AI-related technologies are giving serious ramifications for a wide range of military capabilities, ranging from logistics-support to autonomous weapons systems.

With this understanding of AI, this article explores the most likely impacts of AI on the character of future warfare focusing on how AI will change just about the way we fight in the battlefield, analyze the current status and problems of AI applications in the ROK military, and then suggest several policy alternatives with regard to future development of ROK military’s war-fighting capabilities with AI. What makes this study different from existing studies is not simply to discuss the effects of military applications of AI, but to stress the need to analyze changes in future AI-driven warfare and to prepare necessary measures to effectively support them. In particular, an analysis of future warfare in terms of three aspects discussed below would be significant in the sense that it presents a concise conceptual framework to help discuss AI-driven warfare in the future.

2. Future Warfare

As Becker and DeFoor put it, “envisioning a future war is difficult, particularly as we push beyond a decade”[4]. It is because “war can take many forms”[5]. However as regards future warfare as it is linked to AI and autonomous systems, one thing is clear that over the next 20 years, technological advances in AI and autonomous systems will offer enhanced military capabilities to those states that adopt earnestly and operationalize them.

Within the full spectrum of conflicts ranging from low-intensity to high-intensity, this chapter will explore how advanced military technologies with AI will radically change the way we fight, focusing on the factors and circumstances that advanced military technologies with AI will most directly affect the military operating environment, operational concepts and doctrines, commander’s decisions, and acquisition of weapons systems. This discussion will allow

us to address what the ROK military might need to do in the future and should be, ultimately, preparing the armed forces to stand ready when the unexpected war occurs.

2.1. The Acceleration of the multi-domain battle

Traditionally, each service has its own primary, or specific operational domains(that is, land, sea, air and space). However, “the convergence of dramatic improvements in the fields of robotics, artificial intelligence, materials, additive manufacturing, and nanoenergetics is dramatically changing the character of conflict in all domains” [6]. In this process, as Creviston indicates, “AI...will be best operationalized not by a dedicated Service or force structure...but by their incorporation into the existing forces in all domains” [7].

It is a clear that each service’s current war-fighting capabilities have some limitations to adapt effective operations simultaneously across domains, given that “cyber and space domains may become tomorrow’s most valued battlespace” [8]. Albeit “cyberspace has distinct technologies and methods, it shares many characteristics with the traditional domains” [9]. As a result, it is vital to understand the cyber domain and how it relates to the traditional domains of land, sea, air and space to ensure victory in the future warfare. Indeed each service’s dominance in the traditional domains will be inadequate for coping with the rapidly evolving AI-driven multi-domain operational environment. It is because, as Perkins and Holmes argue, that “as advancements in cyber and electromagnetic spectrum, robotics, artificial intelligence”...and a host of others continue to accelerate and proliferate across multiple domains...we can no longer develop domain-specific solutions that require time and effort to synchronize and federate” [10].

In this context, the key to multi-domain operations is to integrate all battlespaces to create an operational environment that is more favorable than the adversary and to suppress adversary attacks. To this end, Stafford argues that “a common and reliable data network will need to be shared among the Services” [11]. Davis emphasizes the need for the integration of every Service’s data networks as follows. “Being able to rapidly process the flood of information from varied platforms operating in multiple domains translates into two fundamental military advantages - speed and range. Moving faster than your adversary enhances offensive mobility and makes you harder to hit. Striking from farther away similarly benefits the element of surprise and minimizes exposure to enemy fire”...“AI makes it possible to analyze dynamic battlefield conditions in real time and strike quickly and optimally while minimizing risks to one’s own forces” [12]. As a result, regardless of whether each Service wants it or not, rapidly evolving technologies with AI will radically alter the way we fight in the future and will increase multi-domain operation, thereby requiring enhanced multi-domain capabilities, such as new operational concepts and doctrines, and system of systems(ISR+C4I+PGMs).

2.2. The generalization of cognitive-centric operations

With the development of AI technology, many changes will take place in the military operations. Among these changes, the most remarkable thing is that the centerpiece of operations will change from physical domain to non-physical, or intangible(cyber) domain. Indeed, the cyber domain is a unique zone of operations and is already highly militarized. Hence, to dominate the intangible battlefield, it is essential to achieve ‘cognitive dominance,’ or ‘intelligence dominance,’ meaning, the ability to make better and faster decisions than the adversaries through the active use of human-machine interaction relationships.

Kania argues that “the essence of cognitive dominance is to interfere with and control adversary’s mental consciousness, thereby subverting combat styles and will” [13]. In this process why are human-machine interaction relationships, rather than human-specific cognitive ability alone, more needed in conducting operations? It is because human-specific cognitive ability alone cannot match the ‘operational tempo’ that AI singularity offers. Cyberspace is a good example of this. Action in cyberspace happens at speeds beyond human cognition.

In the cognitive-centric operations, the pattern of engagement is expected to shift from a 'destruction-based war-fighting' utilizing kinetic and chemical energies to a 'disability-based war-fighting' utilizing directed and biological energies and software means. Here, the key of disability-based war-fighting is to induce the disturbance of adversary's C2 system and combat system through the various ways of stealth strikes, functional paralysis, and conscious control, etc.

From this point of view, the goal of cognitive-centric operations is to preempt the military decisionmaking cycle called the 'OODA loop'(observe-orient-decide-act) faster than the adversary. It makes the adversary completely unprepared for a defense, thereby ultimately winning without a fighting. As Weinbaum and Shanaban properly describe, "operating inside an adversary's OODA loop helps accomplish those objectives by disorienting or warping an opponent's mental images so that he can neither appreciate nor cope with what is happening around him..." "Counter-AI will become prevalent while influence operations will take on new dimensions that have yet to be fathomed, requiring a renewed emphasis on both offensive and defensive cognitive-centric operations" [14].

Indeed, as Finkel argues, "meeting the challenge of AI singularity...will require counter-measures. it may be worthwhile to consider developing measures to misguide enemy AI learning and decision making through new kinds of deceptions" [15]. Due to this, the use of a high-performance AI system will be inevitable in the process of collecting, processing, and analyzing information, and achieving decision superiority through all-domain C2 to achieve cognitive dominance faster than the adversary. In this respect, adopting the use of AI to achieve decision superiority in the future battlefield will be no longer be an option but a necessary condition for commanders, thereby increasingly relying on AI staffs.

2.3. The acceleration of human-machine fusion and increased ambiguity of blurring distinction between combatants and non-combatants

Supposing that AI will eventually be applied in military affairs, AI-driven autonomous systems are being increasingly used to augment, rather than simply replace, individuals and platforms. In fact, the augmentation of human systems with autonomous systems will permit longer-term mission performance, enable greater lethality, develop various measures to protect physical platforms from adversary's attack, and enhance human and military unit's operational performance.

As a result, the future warfare will lead to new forms of combat operations in terms of human involvement in the execution of autonomous systems' tasks, composing of 'human-in-the-loop system' and 'human-out-of-the-loop system.' Here, in the case of human-out-of-the-system, I believe, only if the required level of trustworthiness in terms of international ethics and legal criteria is achieved can the question of the use of combat operations be addressed by defining its specific role.

Human-in-the-loop system, that is, human-machine C2 relationship, is also ethically and legally responsible for the combat actions [16]. In connection with this matter, as Ray, Forgey, and Mathias rightly indicate, "lethal force can only be committed to a target when appropriate levels of human judgment by an individual have been taken into account. In other words, a person, not a machine, has the ultimate decision making authority"... "Planning and direction will likely continue to be a human-driven operation, although AI/AS can provide recommendations using historical data, cultural knowledge, previous operational design, and result" [17]. Indeed, it is critical to ensure that a final decision is only to be made by a human.

The value and importance of AI technology will depend on how AI technology assists and supports the human. Indeed, although disruptive technologies including AI are doing a major role in conducting warfighting, war still remains a human endeavor. Herr argues that "humans are in no way perfect, of course, but our ability to identify patterns and integrate information holistically is superior to computers in many situations and is a tool that can help maintain

situational awareness. Furthermore, without humans in the loop, it may be difficult for commanders to know when systems have been compromised, as feedback from a compromised system may not accurately represent its status, location, or activities. Humans will not be able to intervene against all types of attacks...but we may be able to intervene against misleading signals from sensors and other challenges” [18].

As a result, the success or failure of AI-driven warfare will depend on the commanders’ stable C2 capability in an extremely complex operational environment, which will increase the importance of ‘hybrid intelligence,’ referring to an intelligence system that connects the human brain directly to a computer and combines human intuition with AI’s data processing and analysis capabilities. In addition, nonphysical aggression in the intangible domain will also spread and thus will encourage a variety of civilian experts to participate in war-fighting. The ratio of proxy wars by non-state groups will also increase. For this reason, it will be now meaningless to distinguish between front and rear, and between wartime military operations and non-wartime military operations. A number of noncombatants will be deployed and operated throughout the entire domains as main actors of war-fighting activities in the future.

3. AI Applications in the ROK Military: Current Status and Problems

3.1. Current status

The ROK military is promoting the military buildup based on the 4th industrial revolution technology as one of the means to build capacity for ‘Defense Reform 2.0’ and ‘Wartime OP-CON transition.’ It is currently pursuing a wide range of reforms, including reorganization of the military structure and military enhancement project in accordance with the Defense Reform 2.0 [19]. In particular, as the operational area of the Army Corps has been greatly expanded, it is faced with the problem of conducting operations with a small number of troops and weapon systems. In order to effectively cope with this, the army is currently making efforts to mobilize infantry units by grafting AI+ICBM(IoT, Xcloud, Big Data, Mobile) of the 4th industrial revolution, and perform multi-domain operations in complex battlefield situations by making all platforms hyper-connected networks. One of the representative programs for achieving this is the ‘Mount Paektu Tiger Project’(MPT), aiming ultimately to connect all unit platforms(warrior platforms, drone bots, combat vehicles, etc.) in both manned and unmanned systems using AI+ICBM(IoT, Xcloud, Big Data, Mobile) to perform multidomain operations. It makes it possible to share information in real time and to enable all ground troops to move by wheeled-armored and other small tactical vehicles, thereby achieving quick and fatal response to various unspecified threats in the future [20].

The Air Force is currently working on a project, so-called ‘Space Odyssey Project.’ It is a sort of plan to apply AI technology to a variety of space power-related weapons systems. It consists of three stages. The first stage, which aims to be achieved by 2030, aims to establish the capabilities necessary to equip the space surveillance-based system, link the missile defense system and the space surveillance system, and strengthen centralized control over the space domain. It is planned to incorporate big data and hyper-connected technologies. The second stage, which is aimed to be achieved by 2040, aims to secure an ability to replace space assets(aerospace power) and to secure the capabilities necessary to equip space resilience by utilizing aviation assets. It is planned to combine AI and unmanned swarming technology. The third stage, aimed at around 2050, is to establish a ground-based anti-space operation system, establish an air-space-based anti-space operation system, and secure the capabilities necessary to insure selective space superiority in the space domain. It is planned to incorporate fully autonomous AI and high-power directed energy technology [21]. Not enough R&D investment, however, has been made in the above-mentioned technologies that can dramatically increase war-fighting capabilities of the Air Force.

The Navy is now working to implement a project, called 'SMART Navy,' which is the blueprint for the future navy. This is a plan to apply information and communication technologies including AI to the naval platforms. Through this, it aims to upgrade the naval weapons systems, intelligentize the C2 system, and strengthen the domestic and international cooperation system. Recently, as part of the SMART navy project, wireless network system was established in Yangmanchunham(DDH-I, 3200t), and the latest ICT technologies including AI technology are applied. This is to show advanced battleships and communication systems that have been in development for years under the project name "Smart Battleship and Operations[22]. In addition, the Navy also has deployed the Super aEgis II machine gun that can identify, track and destroy moving targets without human intervention.

3.2. Problems

The proliferation of AI causes security problems not only in the military field, but also in the whole society. Hence, a comprehensive response at the national level is required. South Korea, however, has not yet developed such a comprehensive level of responsiveness. In fact, mobilizing resources in developing AI and autonomous systems requires political and military leaders' strong leadership. The Blue House, the National Assembly, and the Ministry of National Defense(MND), however, have neither taken a more active approach in developing policies, or strategies to govern AI technology.

One of the most obvious problems related to AI due to the lack of such national-level strategies and guidelines is the shortage of AI experts who are completely focused on that crucial technology. South Korea currently lacks the professional manpower necessary to strengthen AI competitiveness. For example, out of the 24,000 AI professionals working globally, only 1.8% of the workforce is working in South Korea, which is lower than that of India[23]. Due to the rapid growth of the AI industry, the supply and demand of professional manpower itself is unbalanced, and the development of human resources through educational institutions is also progressing slowly in South Korea.

This lack of strategic guidance and funding, as well as the shortage of AI-related professionals, has an equally negative impact on the ROK military application of AI. The first problem that can be pointed out is that the priorities of the military application of AI technology are not clear. There is no clear direction on whether to apply it to logistics-support or cyber fields first, and then to all fields of defense, or to all fields of defense from the beginning. Due to this reason, the current ROK military application of AI is just conceptual and only exists mostly on paper, and few investment have actually been made. Indeed, the government does not act to convert AI technology for military use, and not adopt it for operational concepts and doctrines and acquisition process of weapons systems. As a result, in the case of the ROK military, it will take a considerable amount of time to verify AI-based operational concepts and doctrines and weapons systems through combat experiments at joint, or each service level.

The second problem that can be pointed out is that the number of AI-related military personnel is vastly insufficient compared to the private sector. In some way, this is being a major factor that fails to utilize the full military potential of AI. In such a lack of professional manpower, efforts must be made to cultivate professional manpower through military education and training, but even this is not being done properly. AI-related education and training for officers are currently not properly conducted. It is because there is no course at the National Defense University or each military school to educate officers on how AI technology will affect all areas such as military strategy, operational doctrine, military structure, and acquisition of weapons systems. As a result, it appears that much of the ROK military's discussions on military application of AI remain at a level that mimics the roadmap of the US and other advanced countries.

The third problem is that each service, on the surface, emphasizes the need for military application of AI, but in reality each service is still tied to acquiring legacy weapons systems

like tank, submarine, light aircraft carriers, fighters, etc. It does not require AI-related technologies to be used when improving performance of existing weapons systems or acquiring a new weapons systems. A typical case in point is the cyber field. If AI is used as a means of cyber-attack by non-governmental organizations and companies related to the military, it can become our weakness. However, preparations for countermeasures for this have not yet been sufficiently carried out.

In short, the current AI applications in the ROK military is just conceptual and only exists mostly on paper, and few investments have actually been made. To solve these problems as early as possible, we need to make the following policy efforts.

4. Policy Suggestions

AI is not yet widely used in actual battlefields. It, however, is believed that the potential impact on the way we fight in the future will be huge. For this reason, many countries are actively pursuing various AI-related technology development strategies that can be used militarily. However, at present, South Korea is somewhat inferior in AI competitiveness compared to US, China, and EU in terms of overall national competitiveness. The US is leading rivals in development and use of AI while China is rising quickly and the European Union is lagging[24]. With this in mind, South Korea needs to consider the following measures.

First, it needs to establish an AI Organization under the Blue House or the Office of the Prime Minister, similar to the US's 'National Security Commission on AI' composed of experts recommended by the US Department of Defense and the Senate and House of Representatives. Through this sort of organization, it needs to provide national guidelines and directions for AI-related budget investment and manpower training. It is pointed out that national support should be devoted to fostering professionals who converge and research AI and related industries. In addition, it is necessary to comprehensively respond to new military and security threats imposed by new advanced technologies and to promote international cooperation related to AI ethics[25][26]. In this process, it needs to provide safeguards for AI to be ethically controlled within the military system.

Second, it is necessary to install and operate AI-related departments so that they can play a central role in promoting defense and military-related AI strategies at the level of the Ministry of National Defense(MND). It continuously grasps and analyzes military threats such as the trend of AI-based military power construction in neighboring countries surrounding the Korean peninsula, and suggests strategies and directions for AI that can be applied to the defense and military fields. In addition, the military, universities, research institutes, commercial and defense companies need to play a role of presenting mid/long-term plans necessary for cooperative development of AI-related technologies. MND needs to demand each service to consider at least one AI technology or AI-related alternative whenever acquiring new weapons systems. In this process, it will be important for governments and defense industries to work together to reduce the risks of adopting new war-fighting technologies, while maximizing the benefits that such technological evolution brings.

Third, future warfare posed by AI and autonomous systems is expected to show a big difference from modern warfare in all areas such as military strategy, operational doctrine, organizational structure, logistics-support, and weapons systems. Therefore, the government needs to establish an AI-related response organization, such as 'Artificial Intelligence Center' at the Joint forces' level, and to lead and meticulously analyze the patterns of AI-driven future warfare, and to suggest various ways to respond to this. In particular, it is necessary to focus on developing joint operational concepts and doctrines that can cope with the threats, and developing military-specific 'cloud computing technology' that can handle large-scale data for joint military operations. In this process malicious code or source can be implanted in AI functions in the process of data processing of cloud-based computing, hence it will be necessary

to prepare for cyber-attacks from the enemy. These efforts will serve as a factor that makes military education and training practically possible in a virtual environment applying AI, big data and IoT technologies.

Fourth, it is important that the military not only invest in new high-tech weapons systems, but also help military personnel to adapt to rapidly changing battlefield environment. For this, there is a need to establish an education system that can cultivate basic knowledge for AI-enabled warfare in the National Defense University, the Joint Military University, the Army, the Navy, and the Air Force University. Through this professional education, it will be possible for military officers to understand exactly what role AI plays in combat operations, especially how AI performs in support of commander's decisions.

In fact, as Zais correctly points out, "the importance of AI-related education increases with AI aspirations and the illusion of progress. Without that education, we face a world where senior leaders use AI-enabled technologies to make decisions related to national security without a full grasp of the tools that they -and our adversaries- possess. This would be equivalent to a combat arms officer making strategic military land-power decisions without foundations of military education in maneuver warfare and practical experience" [27]. As a result, a variety of basic and applied courses related to AI-related military strategy, operational concepts and doctrines, and weapons systems should be established so that military officers can cultivate AI-related various knowledge. It is because, as Sipper indicates, "a well-trained and well-educated workforce is...absolutely critical to maintaining security and operations" [28]. In this process, education and training through civil-military collaboration, rather than the military's own education, will be more effective, given that the private sector is more advanced than the military.

Fifth, at the level of land, sea, and air forces, it is necessary to make efforts to identify and acquire the core technologies of AI suitable for the characteristics of self-defense in creating the requirements for weapons systems. Since it is expected that AI-based advanced weapons systems will prevail in the battlefield in the future, the rapid acquisition of AI-based new weapons systems is the key to ensuring the asymmetry of warfare, which will make adversaries to rethink how they project combat power.

These five measures, I think, will define how our own war-fighting capabilities might need to evolve to outpace our competitors.

5. Conclusion

AI is calling for new 'revolution in military affairs'(RMA) in how we approach our national defense and security. Despite concerns over the ethical aspects of AI, the military application of AI will be accelerated, thus making it possible to create new operational concepts and weapons systems necessary for the coming changes in warfare. As time passes by, it will cause a huge shift in the warfare.

It appears that the ROK military is, however, somewhat lagging behind in adapting to the rapidly changing character of warfare that AI singularity offers – compared to the countries surrounding the Korean peninsula, such as China, Russia, and Japan in terms of AI strategies and strengths for military purpose. This is due to our national and military leaders' lack of basic understanding of what AI is and what effects can be achieved in the military realm. The lack of understanding can lead to a miscalculation in the use of AI-driven war-fighting forces and capabilities in execution or support of national security and defense objectives.

Accordingly, the ROK military may need to take more risks in an effort to prepare for the coming changes in warfare, and needs to place a greater focus on AI, calling for significant and urgent changes to operational concept, acquisition and sustainment of weapons systems, etc. Indeed the military needs to discard ineffective legacy weapons systems and prioritize investment of advanced technologies including AI, and streamline its cumbersome acquisition

& sustainment process and procedures as a means to win the AI race with other countries. With AI, the ROK military will be able to create more agile, flexible and lethal war-fighting capabilities in strategic, operational, and tactical dimensions.

6. References

6.1. Journal articles

- [1] Choi J. Current and Future Considerations for the Use of Artificial Intelligence by the United States' Department of Defense. *Robotics & AI Ethics*, 6(1), 1-6 (2021). [\[Article\]](#)
- [3] Pfaff CA. The Ethics of Acquiring Disruptive Technologies: Artificial Intelligence, Autonomous Weapons, and Decision Support Systems. *Prism*, 8(3), 129-129 (2020).
- [4] Becker JJ & Defoor JE. Exploring the Future Operating Environment. *Joint Force Quarterly*, 89(2), 123-123 (2018).
- [5] Hoffman FG. Examining Complex Forms of Conflict: Gray Zone and Hybrid Challenges. *Prism*, 7(4), 32-32 (2018).
- [6] Hammes X. Cheap Technology will Challenge U.S. Tactical Dominance. *Joint Force Quarterly*, 81(2), 76-76 (2016).
- [7] Creviston DD. Transforming DOD for Agile Multidomain Command and Control. *Joint Force Quarterly*, 97(2), 83-83 (2020).
- [8] Woods KM & Greenwood TC. Multidomain Battle: Time for a Campaign of Joint Experimentation. *Joint Force Quarterly*, 88(1), 16-16 (2018).
- [9] Denning DE. Rethinking the Cyber Domain and Deterrence. *Joint Force Quarterly*, 77(2), 8-8 (2015).
- [10] Perkins DG & Holmes JM. Multidomain Battle: Converging Concepts Toward a Joint Solution. *Joint Force Quarterly*, 88(1), 54-54 (2018).
- [11] Stafford JC. Multidomain Ready: How Integrated Air and Missile Defense is Leading the Way. *Joint Force Quarterly*, 100(1), 103-103 (2021).
- [12] Davis Z. Artificial Intelligence on the Battlefield: Implications for Deterrence and Surprise. *Prism*, 8(2), 118-118 (2019).
- [13] Kania EB. Minds at War: China's Pursuit of Military Advantage through Cognitive Science and Biotechnology. *Prism*, 8(3), 87-87 (2020).
- [14] Weinbaum C & Shanahan JN. Intelligence in a Data-driven Age. *Joint Force Quarterly*, 90(3), 6-7 (2018).
- [15] Finkel M. AI Singularity and the Growing Risk of Surprise: Lessons from the IDF's Strategic and Operational Learning Processes, 2014-2019. *Prism*, 8(4), 38-38 (2020).
- [16] Kim H & Park G. Ethical Issues on AI Equipped Combat Robots. *Robotics and AI Ethics*, 5(2), 1-7 (2020). [\[Article\]](#)
- [17] Ray BD & Forgey JF & Mathias BN. Harnessing Artificial Intelligence and Autonomous Systems Across the Seven Joint Functions. *Joint Force Quarterly*, 96(1), 120-120 (2020).
- [18] Herr A. Will Humans Matter in the Wars of 2030?. *Joint Force Quarterly*, 77(2), 78-78 (2015).
- [19] Lee S. The Strategies of ROK's Defense Reform with Special Reference to the Peace Process and Arms Control. *International Journal of Military Affairs*, 5(2), 1-8 (2020). [\[Article\]](#)
- [25] Li Y & Park G. AI Ethics and Privacy Right. *Robotics and AI Ethics*, 5(2), 27-33 (2020). [\[Article\]](#)
- [26] Kim T & Park G & Seo E. IR4.0 and Ethical Tasks of AI. *Robotics and AI Ethics*, 4(2), 6-13 (2019). [\[Article\]](#)
- [27] Zais MM. Artificial Intelligence: A Decision Making Technology. *Joint Force Quarterly*, 99(4), 72-72 (2020).
- [28] Sipper JA. It's not Just about Cyber Anymore: Multidisciplinary Cyber Education and Training Under the New Information Warfare Paradigm. *Joint Force Quarterly*, 100(1), 50-50 (2021).

6.2. Books

- [2] Morgan FE & Boudreaux B & Lohn AJ & Ashby M & Curriden C & Klima K & Grossman D. Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World. RAND (2020).
 [20] ROK Army. ROK Army, towards 2030. Army HQ (2021).
 [21] ROK Air Force. Air Force Quantum 5.0. Air Force HQ (2020).

6.3. Additional references

- [22] <https://www.yna.co.kr/>(2020).
 [23] <https://www.asiae.co.kr/>(2021).
 [24] <https://www.scmp.com/>(2001).

7. Appendix

7.1. Authors contribution

	Initial name	Contribution
Author	JK	<ul style="list-style-type: none"> -Set of concepts <input checked="" type="checkbox"/> -Design <input checked="" type="checkbox"/> -Getting results <input checked="" type="checkbox"/> -Analysis <input checked="" type="checkbox"/> -Make a significant contribution to collection <input checked="" type="checkbox"/> -Final approval of the paper <input checked="" type="checkbox"/> -Corresponding <input checked="" type="checkbox"/> -Play a decisive role in modification <input checked="" type="checkbox"/> -Significant contributions to concepts, designs, practices, analysis and interpretation of data <input checked="" type="checkbox"/> -Participants in Drafting and Revising Papers <input checked="" type="checkbox"/> -Someone who can explain all aspects of the paper <input checked="" type="checkbox"/>

Publisher: J-INSTITUTE
ISSN: 2435-3345

Website: www.j-institute.jp/ai/
Editor: ai@j-institute.jp

Corresponding author
E-mail: allee@cku.ac.kr

dx.doi.org/10.22471/ai.2021.6.2.21

© 2021 J-INSTITUTE

Analyzing the Effects of AI Education Program based on AI Tools

Aeri Lee

Catholic Kwandong University, Gangneung, Republic of Korea

Abstract

Purpose: In recent years, AI technology has emerged as a core technology for the fourth industrial revolution. Along with the existing SW education, it is time to prepare for AI education, since the interest in AI education is also growing. The fact that learning about the theory and principles of AI is important is also stressed in the college education. However, non-major students tend to perceive AI technology to be difficult, and hence, a more efficient AI education is required. The purpose of this research is to guide the learners' direction for AI education, helping them to break away from their fears of AI through the AI education appropriate for the learners' level.

Method: In this study, to analyze the effect of the developed educational program on the learners' AI literacy, the dependent variables changed before and after the application of the program were analyzed, and the single group's ex ante and ex post testings were applied for the designing method for inferring the causal relationship.

Results: The AI literacy of undergraduate students was analyzed for the educational program via the ex ante and ex posts testings of a single group by operating this program. As a result of the analysis, significant results were obtained where all three areas of the AI literacy were improved through the educational program

Conclusion: This study is meaningful in that the AI educational program was proposed for the liberal arts class at college for the non-majors, and that the AI educational program which is generally applicable for the liberal arts class was developed and directly applied to validate its effectiveness. Discussions are expected to continue to develop the various details of the AI basic education to ensure that the non-majors can enter the AI education without fearing the AI based on the results of this study.

[Keywords] Artificial Intelligence, AI Education, Education Model, AI Literacy, Artificial Intelligence Ethics

1. Introduction

The fourth industrial revolution refers to a hyper-connection era where not only humans but also all devices are connected through AI, the Internet of Things, and the Cloud based on the big data. AI is a core technology for the 4th industrial revolution, and the interest in AI has grown even more significantly across various fields utilizing the information generated in a large quantity[1].

AI is now regarded as a core engine which will lead innovative changes in society and economy beyond new technologies, while attracting attention as a crucial and essential tool across almost all areas of the modern society. Hence, many countries around the globe, including the United States and China, are devising strategies at the national level, and are actively investing to secure and foster AI talents, which are a key element of the AI technology's competitiveness[2][3].

The United States government is strengthening the STEAM education as a universal education for all citizens to enhance the AI capacity, while reinforcing the AI related education across the

elementary, middle, and high school education. The Chinese government has also introduced the AI education at elementary and secondary schools to develop the curriculum and textbooks, while strengthening the teachers' education. Furthermore, they are endeavoring to foster talents by reorganizing the college education system and opening the AI related departments. The Japanese government has also introduced the "information" curriculum for the AI knowledge education and college entrance examinations for the elementary, middle and high schools. Moreover, colleges and universities are developing an AI standard curriculum and making efforts to foster professional AI talents[4]. Accordingly, the Korean government began preparing for the AI education by implementing the AI education across all occupational groups for lifetime and by announcing the national AI strategy with a view to foster the world's finest AI talents[5].

Consequently, it may be known that, the fostering of AI talents at each country around the globe operates in two directions, such as the process for fostering professional talents related to AI, and the approach for universal education which anyone may know and utilize concerning AI. Hence, beyond the AI major education, AI education is required as a universal education for the non-major students.

However, since the existing AI related policies are focused on the universal education for the elementary and secondary schools and the major education at colleges, the AI related basic knowledge and education for the liberal arts studies at college for universal education are not adequate[6]. As for the artificial intelligence technology, it is not easy to learn it over such a short period time or achieve proficiency. Furthermore, many difficulties present themselves in developing educational programs by structuring the industrially concentrated artificial intelligence technologies for education purposes. In particular, the need for an education optimized for learning to understand artificial intelligence is very large, yet the research on the AI curriculum as a liberal arts education for college students with little artificial intelligence related experiences and knowledge is significantly inadequate, and hence, the development of an educational program in line with the learners' needs at their eye level is urgent[7]. In such an environment, it is not easy for the students to find answers as to what they need to learn about AI. Furthermore, while they are aware that it is necessary to deal with the AI related areas in their fields of major, they fear that AI is "difficult." For such reasons, it is not easy for the non-majors to expand their fields of study by making convergence with AI.

Accordingly, this study was carried out with a view to provide guidance for the learners' direction for the AI education, breaking away from fears of AI through the AI education appropriate for the learners' level. Towards this end, after operating the AI tool-based AI education program which may be utilized by non-majors, it is intended to present a direction for education for AI via the validation of its effectiveness by analyzing the learners' AI literacy. The program proposed in this study is expected to be utilized as an effective program to help improve the AI literacy for the liberal arts education at colleges and universities moving forward.

2. Theoretical Background

2.1. AI education

AI is regarded as a core technology of the fourth industrial revolution, which is called the industrial revolution for the next generation, while the ability to understand and utilize AI is pronounced as a core competency and an essential literacy required for living in and pursuing the future society[8].

AI education is conducted as an extension of the SW education initiated in the United Kingdom in 2014, and many countries are preparing policies to implement the AI education for the elementary, middle, and secondary school education in the recent few years[9].

Wong[10] presents the application of the concept of AI as a solution for real life problems as an important goal of achievement, and also proposed an approach of problem-based learning

for the AI education. Furthermore, the AI education for elementary, middle, and high schools is generally conducted as an extension of the existing software education and information education, while the problem solving experiences utilizing AI are stressed.

Kim S et al.[9] stressed that the AI education for the elementary, middle, and secondary school is conducted as an extension of the SW education, and as for the AI education, AI literacy for the AI users must be included, and the capability of achieving creative products along with the problem solving skills applying AI technology based on the current computing thinking skills and coding education capacity. Hence, the AI education ought to provide education, focusing on the problem solving experiences addressing problem situations in everyday life such as the SW education.

The discussion of how to commence the AI education has just begun. Various attempts are made domestically and internationally toward finding answers rather than presenting answers thereto. In Korea, too, some studies have been attempted to conduct the AI education and analyze it[11].

Kim K & Park Y[12] presented and applied a step-by-step model of the AI education for the elementary school students, while[13] confirmed changes in the attitudes towards the AI technology of the elementary school students after conducting the AI education based on the block programming language for the elementary school students.

The AI related studies for the undergraduate students for the liberal arts education in college is inadequate relative to the studies for the elementary, middle and high school students. Oh K & Kim H[4] ran a course for liberal arts education for 1 credit for 5 weeks so that the learners interested in AI may learn about big data correlated to AI, after which analyzed the learners' attitudes towards AI and validate its effectiveness. As for Woo H[14], most of the AI related academic subjects are structured around undergraduates studies related to computer science, the need was proposed to operate the AI based academic subjects of liberal arts education which the non-majors may also access. Jun S[6] developed an AI education program based on the experiential learning for the AI liberal arts education for the non-major students.

In such a situation where the attempts for AI education are made, this study intends to develop and apply an education program which provides the experiences for the students experiencing the AI education for the first time to understand the concept of AI and explore and solve AI problems, following which to analyze the results of the application.

2.2. AI literacy

Literacy is defined as literacy, whose definition is expanding in line with the globalization and diversification of the society, including the utilization of what one understands to create products and participate in and engage in activities in the society. AI literacy is a concept which was created by the development of the AI technology, and is also a capability required for the future society where AI essentially coexists and changes[15]. Long D & Magerko B[16] is defined as the capability for individuals to critically evaluate and use the AI technology, and communicate and cooperate effectively with the AI to utilize the AI. Furthermore, it was claimed that the AI literacy ought to include the computational literacy, scientific literacy, and data literacy, including the digital literacy. Wong[10] provided the concept of AI, AI application, AI ethics and safety as the sub-elements of the AI literacy. Lee C[17] explained it as the ability to understand the concept of AI, use AI as a tool, and use the AI technology to solve problems to produce outcomes, while presenting the AI related basic knowledge, AI utilization ability, AI development ability, and AI ethics and values[18][19][20] as the sub-elements as illustrated in <Table 1>.

Table 1. Sub-elements of AI literacy.

Sub-elements	Contents
AI related basic knowledge	Ability to explain the concept and functions, etc., of core technologies such as the concept of AI and machine learning.

AI utilization ability	Ability to solve problems by applying AI tools to one's own life.
AI development ability	Ability to use the AI technology and data to produce outcomes such as learning models required for the problem solving.
AI ethics and values	Attitude of taking responsibility for the results, while upholding moral values and ethics in and for the utilization and development of the AI technology.

2.3. AI tools

Various educational services for experiencing and learning about AI have recently been developed and utilized. The variety spans from the kits with which one can make AI speakers seen familiar in daily life, to the tools with which one can implement the core principles of machine learning through programming.

The tool with which one may learn about machine learning, which is the core of AI, can link the machine learning model implemented as a tool which may be experienced on the web via various examples of machine learning principles with programming.

ML4KID(Machine learning for kids) was developed by Dale Lane, the developer, to teach students the concept of AI and how to use it. Using the API of the IBM Watson Developer Cloud, a text / image / number / sound learning model can be made and programmed by linking with Scratch, Python, and App Inventor[21].

Teachable Machine is a website provided by Google, where neural network among machine learning algorithms can be learned, and where the experiments for verifying the results can be conducted. TensorFlow API, Google's machine learning tool, is utilized. Teachable Machine can create a learning model by using image / audio / pose data[22].

Cognimates is a service provided by the MIT Media Lab which developed Scratch. It is a service which can be directly modeled as an AI Literacy education platform for the beginner learners and can also be applied to the user projects. For the starter, a starter kit and pretrained model are provided[23].

Such AI tools are also very useful not only for the elementary, middle, and high school students who need to learn about the characteristics of AI, but also for the general public and non-majors. This is because it offers the scalability and applicability which may be applied to the people of various ages or background knowledge. In the AI design process, data and algorithms may be transformed into a variety to understand how the elements of AI are involved. The advantages of utilizing the AI educational tools in the classroom are as follows.

First, no professional computing knowledge or terminologies are required. Second, it is possible to utilize text, numbers, and sound data as well as images for data input. Third, in the latter part of the utilization, coding tools such as Scratch and Python may be chosen as per one's own background knowledge, and the AI models reflective of AI ethics may be created.

3. Design of the AI Tool-Based AI Education Program

3.1. Research subject

This study was conducted with 45 students taking "Creative Thinking and Coding," a liberal arts course offered at University C located in Gangwon-do. They were all non-majors, consisted of 29 male students and 16 female students. The "Creative Thinking and Coding" course is a practical class consisted of the computational thinking based coding education and AI curriculum.

3.2. Research design and procedures

In this study, an AI education program which may be utilized by non-majors was designed by using the AI tool. The development of a program for the AI education was intended for the

liberal arts classes for the college students not majoring in computer science. In this study, to analyze the effect of the developed educational program on the learners' AI literacy, the dependent variables changed before and after the application of the program were analyzed, and the single group's ex ante and ex post testings were applied for the designing method for inferring the causal relationship. The performance analysis of the educational program proposed for learning artificial intelligence was performed via a questionnaire for the learners.

The study was consisted of the procedures of 1)the implementation of a preliminary investigation, 2)implementation of 15 sessions of the educational program, and 3)the ex post investigation for the ex ante and ex post analysis of the AI literacy within the single group(N=45). At which time, the AI subject class was conducted as an independent variable.

This study was applied to the "Creative Thinking and Coding" course opened in the Faculty of Liberal Arts at University C of the Republic of Korea. The subjects are 45 people, and this is a class where various students majoring in humanities and social sciences participate. The weekly educational program was operated as illustrated in <Table 2>, and was consisted of the general coding education and the AI curriculum.

Table 2. AI program class contents.

Week	Topic	Key activity
1	Course overview	Computational thinking and AI
2-4	EPL grammar	Computational thinking and programming
5	Understanding AI and exploring problems	Understanding the basic concepts and implementation principles of AI exploring various cases of AI nearby identifying issues and limitations of AI technology
6	Understanding AI technology	Understanding AI, datasets, AI algorithms, and machine learning
7	Practicing AI	Understanding algorithm / data bias AI tools and datasets
8	Midterm exam	
9	Practicing AI	Creation of AI tool and AI model
10	Practicing AI	AI natural language processing project
11	Practicing AI	AI image recognition project
12	Practicing AI	AI number recognition project, decision making tree
13	Developing AI	AI speech recognition project

14	Developing AI	Final project
----	---------------	---------------

From Week 1 through Week 4, students acquire the basic grammar and functions of the educational programming language, and also spend the time to code on common topics on their own. Week 5 and Week 6 are devoted to understanding the basic concept and characteristics of AI, exploring various cases of AI nearby, and identifying the issues and limitations of AI technology through them. Week 7 and Week 9 are devoted to using the AI tool to see the actual examples of the data / algorithm bias and also becoming aware of the importance of datasets. During Week 10 through Week 13, students practice or experience the AI through design, development, implementation, and simulation, etc. The learners collect and prepare the data. The datasets prepared in this manner are entered via the AI tool, which is trained to create the intended AI model on their own. Furthermore, the AI model's performance is evaluated, and if the evaluation results are inadequate, the data are added, and the AI model construction is repeated. The AI model created in this manner is connected to the individual learners' projects to generate various products. In this process, the data and labels may be modified in various versions, and the relevant results may be calculated and compared. Hence, through them, students may directly experience the specific terms and conditions under which the data bias and algorithm bias occur, and also experience key ethical issues of the AI.

3.3. Testing tool

In this study, to verify the changes in the learners' AI literacy, the survey questionnaires were structured as illustrated in <Table 3>. As for the testing tool for the AI literacy in this study, the survey questionnaires of [13] and [14] which analyzed the effectiveness of the AI education program were referenced. As per the definition and sub-elements of the AI literacy, "Understanding AI," "AI Ethics," and "Efficacy Towards the Utilization of AI" were used as the measure of change for the AI literacy, used by restructuring to meet the level of undergraduate students, and measured on a Likert 5-point scale.

Table 3. Composition of questionnaire.

Item	Details	Number
Understanding AI	Concept of AI	1, 4
	Characteristics of AI	2, 3
	Cases of AI	5, 6
AI ethics	Transparency and explicability	7, 8
	Accountability	9, 10
	Safety and reliability	11, 12
Efficacy towards the utilization of AI	Finding problems	13
	Solving problems	14
	Confidence	15

4. Results of the Study

To measure the effectiveness of the AI education program, a corresponding sample t-test was conducted to compare the average of the AI literacy before and after class. SPSS 21 was utilized for the analysis. The performance analysis of the learning program proposed for the AI learning was analyzed via the survey questionnaires for the learners. The survey questionnaires were

evaluated 15 questions utilizing a Likert 5-point scale, and a total of 45 learners answered 15 questions.

“Understanding AI,” “AI Ethics,” and “Efficacy Towards the Utilization of AI” were used for the measure of change for the literacy. <Table 4> illustrates the results of changes for the AI literacy.

Table 4. AI literacy analysis result.

Factor		N	Mean	SD	t
Understanding AI	Pre	45	2.098	.2179	-43.858***
	Post	45	4.016	.1821	
AI ethics	Pre	45	2.084	.2184	-47.454***
	Post	45	4.031	.1781	
Efficacy towards the utilization of AI	Pre	45	2.031	.1975	-49.625***
	Post	45	4.011	.1627	

4.1. Understanding AI

Before class, 2.089 was taken for measurement, and after class, 4.016 was taken for measurement. It seems that before class, the learners determined that the level of understanding for the AI is low. The ex post analytical results were improved to 4.016, and the average difference was statistically significant with a significant probability of 0.000 for all items, respectively($t=-43.858$, $***p<.001$).

4.2. AI ethics

Before class, 2.084 was taken for measurement, and after class, 4.031 was taken for measurement. It seems that before class, the learners determined that the level of understanding for the AI ethics is low. The ex post analytical results were improved to 4.031, and the average difference was statistically significant with a significant probability of 0.000 for all items, respectively($t=-47.454$, $***p<.001$). Through which, it may be inferred that the learners have understood the concept and cases of the AI ethics.

4.3. Efficacy towards the utilization of AI

Before class, 2.031 was taken for measurement, and after class, 4.011 was taken for measurement. It seems that before class, the learners determined that the efficacy for the utilization of the AI is low. The ex post analytical results were improved to 4.011, and the average difference was statistically significant with a significant probability of 0.000 for all items, respectively($t=-49.625$, $***p<.001$).

5. Conclusion

This study was conducted with a view to provide guidance for the learners’ direction for the AI education, breaking away from fears of AI through the AI education appropriate for the learners’ level. Towards this end, after operating the AI tool-based AI education program which may be utilized by non-majors, the relevant educational program was applied and its effect was analyzed in the liberal arts class targeting 45 non-major students.

According to the analytical results of this study, the learners' AI literacy improved significantly after this program was applied. In the survey questionnaires for “Understanding AI,” “AI Ethics,” and “Efficacy Towards the Utilization of AI,” an average of 4 or higher was achieved, and through

the learners' reaction, it was verified that the improvement of the AI literacy and the educational program demonstrated a significant connection. Through the AI education program, the learners strengthened the concept of AI and further deepened their understanding of it, while establishing the concept and need for the AI ethics. The AI education is not merely about memorizing knowledge, and through the practice by which one could obtain specific results, the learners' efficacy towards the utilization of AI could be improved.

If the non-major students were provided with an AI education learning environment via a web based AI tool, the following effects may also be expected.

First, the AI technology may be used relatively easily, and the scope of implementation for the outcome may be further expanded. Second, it is possible to know and learn about the concept, role, and effect of the transfer learning which compels re-learning appropriate for one's own learning data by using the models learned by others. Third, one may experience key ethical issues of the AI by implementing and using the products including the AI functions.

This study is meaningful in that the AI educational program was proposed for the liberal arts class at college for the non-majors, and that the AI educational program which is generally applicable for the liberal arts class was developed and directly applied to validate its effectiveness. Discussions are expected to continue to develop the various details of the AI basic education to ensure that the non-majors can enter the AI education without fearing the AI based on the results of this study.

As for the future tasks of studies, researching ways to systematize the AI educational model via the application of the models of different languages and levels and the analysis of the learners' factors.

6. References

6.1. Journal articles

- [1] Park H. Trend Analysis of Korea Papers in the Fields of Artificial Intelligence, Machine Learning and Deep Learning. *Korea Information Electronic Communication Technology*, 13(4), 283-292 (2020).
- [2] Han J. Changes in Attitudes and Efficacy of AI Learners according to the Level of Programming Skill and Project Interest in AI Project. *Journal of the Korean Association of Information Education*, 24(4), 391-400 (2020).
- [3] Kim Y & Park G. Rubrics and Schoolwide Approach to the Character Education and Some Implications to AI-based Character Education. *Robotics & AI Ethics*, 5(2), 19-26 (2020). [\[Article\]](#)
- [4] Oh K & Kim H. An Analysis of the Influence Big data Analysis-based AI Education on Affective Attitude towards Artificial Intelligence. *Journal of the Korean Association of Information Education*, 24(5), 463-471 (2020).
- [6] Jun S. Development of Artificial Intelligence Education Program based on Experiential Learning for Liberal Art Education. *The Journal of Korean Association of Computer Education*, 24(2), 63-73 (2021).
- [7] Yi Y & Park Y. Establishing a Definition of AI Literacy and Designing a Liberal Arts Education Program. *The Journal of Language & Literature*, 85, 451-474 (2021).
- [9] Kim S & Kim S & Lee M & Kim H. Review on Artificial Intelligence Education for K-12 Students and Teachers. *The Journal of Korean Association of Computer Education*, 23(4), 1-11 (2020).
- [10] Wong G & Ma X & Dillenbourg P & Huan J. Broadening Artificial Intelligence Education in K-12: Where to Start?. *ACM Inroads*, 11(1), 20-29 (2020).
- [11] Lee S. Analyzing the Effects of Artificial Intelligence (AI) Education Program based on Design Thinking Process. *The Journal of Korean Association of Computer Education*, 23(4), 49-59 (2020).
- [12] Kim K & Park Y. A Development and Application of the Teaching and Learning Model of Artificial Intelligence Education for Elementary Students. *Journal of the Korean Association of Information Education*, 21(1), 137-147 (2017).

- [13] Lee Y. An Analysis of the Influence of Block-type Programming Language-based Artificial Intelligence Education on the Learner's Attitude in Artificial Intelligence. *Journal of the Korean Association of Information Education*, 23(2), 189-196 (2019).
- [14] Woo H & Lee H & Kim J & Lee W. Analysis of Artificial Intelligence Curriculum of SW Universities. *The Journal of Korean Association of Computer Education*, 23(2), 13-20 (2020).
- [15] Choi J. Instructor Competency for Innovative Teaching Methods in the Untact Era. *Robotics & AI Ethics*, 5(2), 50-56 (2020). [Article]
- [17] Lee C. Direction of Software Education in Practical Arts for Cultivating Competencies in the AI Era. *The Journal of Korean Practical Arts Education*, 26(2), 41-64 (2020).
- [18] Yi Li & Park G. AI Ethics and Privacy Right. *Robotics & AI Ethics*, 5(2), 27-33 (2020). [Article]
- [19] Kim T & Park G. Future Oriented Smart Society and Ethical Issues. *Robotics & AI Ethics*, 6(1), 35-44 (2021). [Article]
- [20] Kim T & Park G & Seo E. IR4.0 and Ethical Tasks of AI. *Robotics & AI Ethics*, 4(2), 6-13 (2019). [Article]

6.2. Conference proceedings

- [8] Vazhayil A & Shetty R & Bhavani R & Akshay N. Focusing on Teacher Education to Introduce AI in Schools: Perspectives and Illustrative Findings. 2019 IEEE 10th International Conference on Technology for Education (2019).
- [16] Long D & Magerko B. What is AI Literacy? Competencies and Design Considerations. 2020 CHI Conference on Human Factors in Computing Systems (2020).

6.3. Additional references

- [5] <https://www.korea.kr/> (2019).
- [21] <https://machinelearningforkids.co.uk/> (2021).
- [22] <https://teachablemachine.withgoogle.com/> (2021).
- [23] <http://cognimates.me/> (2021).

7. Appendix

7.1. Authors contribution

	Initial name	Contribution
Author	AL	-Set of concepts <input checked="" type="checkbox"/> -Design <input checked="" type="checkbox"/> -Getting results <input checked="" type="checkbox"/> -Analysis <input checked="" type="checkbox"/> -Make a significant contribution to collection <input checked="" type="checkbox"/> -Final approval of the paper <input checked="" type="checkbox"/> -Corresponding <input checked="" type="checkbox"/> -Play a decisive role in modification <input checked="" type="checkbox"/> -Significant contributions to concepts, designs, practices, analysis and interpretation of data <input checked="" type="checkbox"/> -Participants in Drafting and Revising Papers <input checked="" type="checkbox"/> -Someone who can explain all aspects of the paper <input checked="" type="checkbox"/>

Publisher: J-INSTITUTE
ISSN: 2435-3345

Website: www.j-institute.jp/ai/
Editor: ai@j-institute.jp

Corresponding author
E-mail: poshop99@naver.com

dx.doi.org/10.22471/ai.2021.6.2.30

© 2021 J-INSTITUTE

Mediating Effect of the Involvement with Beauty Content in the Relationship Between the Characteristics of YouTube Beauty Content That Uses AI and the Intention to Purchase Cosmetics: Comparison Between Men and Women from 20 to 40 Years Old

Hyangmee Woo¹

Seokyeong University, Seoul, Republic of Korea

Eunsook Kim^{2*}

Seokyeong University, Seoul, Republic of Korea

Abstract

Purpose: This study aims to analyze the relationship between the characteristics of YouTube beauty content and the intention to purchase cosmetics for men and women from 20 to 40 years old. In particular, it aims to present how to utilize YouTube beauty content by gender by analyzing the mediating effect of beauty content involvement centered on male-female comparisons and studying how to use big data in popular videos.

Method: This study was conducted in 17 cities and provinces nationwide with 201 men and 201 women from 20 to 40 years old using SPSS 25.0 as an empirical analysis method for 402 copies of the online questionnaire. In order to verify the mediating effect, it was re-verified through regression analysis and Sobel-test according to the mediated effect verification procedure proposed by Baron & Kenny(1986).

Results: In Question 1, the characteristics of YouTube beauty contents show a significant positive(+) effect on the purchase intention of cosmetics; in Question 2, the characteristics of YouTube beauty content were found to have a significant positive influence on beauty content; while in Question 3, the path that influenced the purchase intention of cosmetics by mediating the degree of involvement in beauty contents was revealed in part; and finally, in Question 4, gender differences were revealed.

Conclusion: It was found that the characteristics of YouTube beauty content and the degree of involvement have a positive effect on the consumer's intention to purchase cosmetics, and there are gender differences. In particular, entertainment and information had the greatest influence on YouTube beauty content characteristics and beauty content involvement. Therefore, to use YouTube as a marketing method in the future, it will be necessary to target male and female YouTube beauty content viewers based on big data and create beauty content that provides both interest and information at the same time.

[Keywords] YouTube Content, Intention to Purchase Cosmetics, Beauty Content Involvement, Entertainment, Big Data

1. The Needs and Purpose of the Study

YouTube, which started its service in 2005, is a platform where anyone can watch and upload videos[1]. and it is the most prevalent and rapidly growing media platform in the world[2]. Since Google acquired YouTube, YouTube has reached 800 million users per month and more than 4 billion videos per day[3]. In Korea, the number of YouTube app users reached 43.19 million as of October 2020[4]. As can be implied from the current status of users, various contents such as hobbies, public relations, daily-life, and broadcasting are shared on YouTube, by individuals and companies alike. In particular, in relation to the media commerce business that has emerged as a new revenue model, beauty-related content is attracting the highest interests of consumers, so SNS channels such as YouTube are

in the spotlight by consumers who want to care for themselves[5]. Companies are also launching products using the names of famous beauty YouTubers for more effective marketing[6].

Therefore, in this study, we want to check whether the characteristics of YouTube beauty contents affect the cosmetics purchase intentions of consumers from 20 to 40 years old who are interested in appearance. In this process, through the use of big data of high views and popular beauty videos in YouTube, we will verify the medium effect of beauty content involvement in regards to taking care of one's appearance by using products that consumers have encountered. Finally, the purpose of this research is to present implications for how to use YouTube beauty contents by studying the above paths according to gender.

2. Theoretical Background

2.1. Youtube beauty content

This is defined in the standard Korean dictionary of the National Institute of Korean Language as the digital production, processing and distribution of various information provided through the Internet or computer communication, text, sign, voice, audio, image, and video[7]. These contents will be described using narration, video, etc. to suit the purpose of production[8]. Depending on the researcher, the content properties are composed in a variety of ways, but in relation to YouTube content, Lee Jin-hee, Kang Myung-bo, and Jeong Yeon-dong saw that there are various properties of YouTube content that can induce users' behavior[9]. In this study, the characteristics of beauty YouTube contents were composed of informativity, entertainment, novelty, vitality, and interactivity.

Informativity is to obtain useful information through a service or product[10]. In the 2020 Internet survey conducted by the Korea Intelligent Information Society Promotion Agency, data and information acquisition for the purpose of using the Internet showed a high rate of 94.8%[11]. Entertainment is a characteristic for recharging and provides an opportunity for relaxation[12]. Novelty is a concept that is mainly used in the tourism field, and the dictionary defines it as something that is original, new and fresh[13]. Vitality is one of the characteristics of information and has an important influence on how consumers evaluate information. Depending on how rich the information presented in the Internet environment is, there may be differences in the ability of consumers to process it[14]. Interactivity is a noun that refers to 'intercommunication and participation' and is a combination of 'inter' meaning 'mutual' and 'active' meaning 'motion and action'[15]. This concept began to be studied in earnest in the media field with the advent of the web and new media[16].

2.2. Beauty content involvement

Involvement is an important concept discussed in connection with consumer's purchasing behavior and information processing process in advertising research[17]. Beauty involvement refers to the degree to which one manages one's appearance by using various tools such as cosmetics, hairstyles, and products, as well as physical beauty[18]. By applying this to Beauty YouTube, the degree of involvement in beauty content was defined as the level of trying to manage one's appearance by using the products encountered on Beauty YouTube.

2.3. Intention to purchase cosmetics

Purchase intention arises when the products encountered through beauty products satisfy the desire to pursue your own beauty and the desire to express yourself through image[19][20]. Because purchase intention represents the possibility of actual purchase behavior[21], this is a very important evaluation item in the marketing field.

3. Research Method

3.1. Data collection

This study was conducted from February 22 to February 25, 2021 with 201 men and 201 women from 20 to 40 years old to analyze the mediating effect of beauty content involvement in the relationship between beauty YouTube content characteristics and cosmetic purchase intentions. Until now, online questionnaires and consent forms were distributed and collected, and 402 copies were analyzed, excluding unfaithful responses and missing values.

3.2. Research problem

H1, YouTube beauty content characteristics will affect the intention to purchase cosmetics.

H2, YouTube beauty content characteristics will affect the degree of involvement in beauty content.

H3, when the characteristics of YouTube beauty content affect the intention to purchase cosmetics, the involvement of beauty content will play a mediating role.

The relationship between H4, YouTube beauty content characteristics, beauty content involvement, and cosmetic purchase intention will appear differently depending on gender.

3.3. Survey method

This study used the online questionnaire method as a survey tool to empirically analyze the research problem. The questionnaire is made of questions divided as follows: 7 on general characteristics and 27 on Youtube beauty content[22], 5 on beauty content involvement[22] and purchase of cosmetics, for a total of 43 questions, including 4 intentional questions[23], based on the Likert 5-point scale.

3.4. Data processing method

The following statistical analysis was performed using the SPSS 25.0 program for the data of this study.

First, to analyze the validity of the measurement tool exploratory factor analysis was performed, and the reliability of the items constituting the factor was analyzed using the Cronbach' s alpha coefficient .

Second, to understand the general characteristics of the research subject frequency analysis was performed, and to understand the level of research variables descriptive statistics analysis was applied.

Third, to find out the correlation between the research variables Pearson' s correlation was performed.

Fourth, to verify the mediating effect of beauty content involvement in the relationship between YouTube beauty content characteristics and purchase intention, a regression analysis was performed[24], and re-verified through Sobel-test according to the mediating effect verification procedure proposed by Baron & Kenny(1986).

The statistical analysis was performed on the basis of the significance level of 5%.

4. Research Results

4.1. Reliability and validity verification

To verify the validity of the measurement tool used in this study an exploratory factor analysis was conducted. Among the factor analysis methods, a principal component analysis was used to minimize

information loss while extracting factors that explain as much of the variance of the original variables as possible, and factors were analyzed using Varimax rotation until their structure was most pronounced while maintaining their independence. Factor classification was composed of one factor when the eigen value was 1 or more, and when the factor loading exceeded .40, it was classified as a corresponding factor(Song, 2016). As for the characteristics of YouTube content, 3 items(No. 20, No. 21, No. 26) that hinder the validity were excluded, beauty content involvement was derived from 5 questions without exception, and cosmetics purchase intent was derived from 4 questions without any excluded questions.

Next, the reliability analysis using Cronbach's alpha coefficient showed that the reliability coefficient of all variables was 0.7 or higher, indicating good reliability.

4.2. Technical statistics

The mean and standard deviation were calculated to determine the level of the research variables measured in this study. The average of YouTube content characteristics was 4.03 for informativity, 3.35 for entertainment, 3.34 for interactivity, 3.49 for novelty, 3.51 for vitality in a scale of a maximum of 5 points; and the average of cosmetic purchase intentions was 3.47 out of 5 points. Additionally, Skewness and Kurtosis were calculated to determine whether the variables satisfy the normality assumption. If skewness is less than the absolute value of 3 and the kurtosis is less than the absolute value of 10, it is judged to be approximated to the normal distribution(Kline, 2005)[25]. All variables were found to satisfy the normality assumption, as shown in <Table 1>.

Table 1. Technical statistics.

Variable		Minimum value	Maximum value	Average	Standard deviation	Skewness	Kurtosis
Youtube content characteristics	Informativity	1.67.	5.00.	4.03.	0.52.	-0.52.	1.20.
	Entertainment	1.20.	5.00.	3.35.	0.64.	-0.08.	0.03.
	Interactivity	1.00.	5.00.	3.34.	0.61.	-0.12.	0.24.
	Novelty	1.00.	5.00.	3.49.	0.69.	-0.15.	0.26.
	Vitality	1.50.	5.00.	3.51.	0.63.	-0.35.	0.28.
Beauty content involvement		1.00.	5.00.	3.44.	0.65.	-0.40.	0.53.
Intention to purchase cosmetics		1.00.	5.00.	3.47.	0.73.	-0.69.	0.63.

4.3. Correlation analysis

Pearson's correlation analysis was performed to find out the correlation between the variables in this study. Cosmetic purchase intention showed a significant positive(+) correlation with informativity($r=.399$, $p<.001$), entertainment($r=.540$, $p<.001$), interactivity($r=.332$, $p<.001$), novelty($r=.384$, $p<.001$) and vitality($r=.497$, $p<.001$) of the characteristics of YouTube content; as well as a significant positive(+) relationship with beauty content involvement($r=.648$, $p<.001$). The degree of involvement in beauty content showed a significant positive correlation with informativity($r=.386$, $p<.001$), entertainment($r=.594$, $p<.001$), interactivity($r=.411$, $p<.001$), novelty($r=.409$, $p<.001$) and vitality($r=.496$, $p<.001$) of the characteristics of YouTube content. Also, since the absolute value of the correlation coefficient between the measured variables was less than .80, the problem of multicollinearity did not appear, as shown in <Table 2>.

Table 2. Correlation analysis.

	Informativity	Entertainment	Interactivity	Novelty	Lively	Beauty content involvement	Intention to purchase cosmetics
Informativity	1.						
Entertainment	.353***	1.					
Interactivity	.325***	.457***	1.				
Novelty	.380***	.524***	.494***	1.			
Lively	.423***	.539***	.459***	.545***	1.		
Beauty content involvement	.386***	.594***	.411***	.409***	.496***	1.	
Intention to purchase cosmetics	.399***	.540***	.332***	.384***	.497***	.648***	1.

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

4.4. Verification of the mediating effect

4.4.1. The mediating effect of the involvement with beauty content in the relationship between the characteristics of youtube beauty content that uses AI and the intention to purchase cosmetics

The regression model was suitable for Hypothesis 1 that was that beauty YouTube content characteristics would affect the intention of purchasing cosmetics ($F=47.774$, $p < .001$), and entertainment ($\beta=.348$, $p < .001$), vitality ($\beta=.228$, $p < .001$), and informativity ($\beta=.175$, $p < .001$) were partly selected in the order of significant influence. The regression model was suitable for Hypothesis 2 that was that YouTube beauty content characteristics would affect beauty content involvement ($F=57.579$, $p < .001$), and entertainment ($\beta=.406$, $p < .001$), vitality ($\beta=.173$, $p < .001$), informativity ($\beta=.136$, $p < .001$), interactivity ($\beta=.103$, $p < .05$) were partially selected in order of significant influence as shown in <Table 3>.

Table 3. Regression verification result of beauty youtube content characteristics on cosmetic purchase intention.

Step	B	SE	β	t	p	F(R ²)
Hypothesis 1 (independence → dependence)						
Informativity → intention to purchase cosmetics	0.248.	0.064.	0.175.	3.879***	<.001	47.774*** (.376)

Entertainment → intention to purchase cosmetics	0.398.	0.058.	0.348.	6.829***	<.001	
Interactivity → intention to purchase cosmetics	0.009.	0.059.	0.008.	0.159.	.874	
Novelty → intention to purchase cosmetics	0.007.	0.055.	0.007.	0.132.	.895	
Vitality → intention to purchase cosmetics	0.264.	0.061.	0.228.	4.324***	<.001	
Hypothesis 2 (independence→mediation)						
Informationality→ beauty content involvement	0.172.	0.055.	0.136.	3.140***	.002	
Entertainment→ beauty content involvement	0.411.	0.050.	0.406.	8.258***	<.001	
Interactivity → beauty content involvement	0.110.	0.050.	0.103.	2.200*	.028	57.579*** (.421)
Novelty → beauty content involvement	-0.001.	0.047.	-0.001.	-0.012.	.990	
Vitality → beauty content involvement	0.177.	0.052.	0.173.	3.399***	<.001	

Note: *p<.05, **p<.01, ***p<.001.

Next, as a result of analyzing the mediating effect of beauty content involvement, the regression model was suitable ($F=62.798$, $p<.001$), and the influence of beauty content involvement on cosmetic purchase intention was significant ($\beta=.440$, $p<.001$). The significant influence of entertainment ($\beta=.170$, $p<.001$), vitality ($\beta=.152$, $p<.001$) and information ($\beta=.115$, $p<.001$) on hypotheses 1 and 2 was partially adopted. In other words, among the characteristics of beauty YouTube contents, entertainment, vitality and information can be said to have an influence on the intention to purchase cosmetics by partially mediating the degree of involvement in beauty contents, as shown in <Table 4>.

Table 4. Results of verification of the mediating effect of beauty content involvement.

Step	B	SE	β	t	p	F(R ²)
Hypothesis 3 (independ →dependence)						
Informativity → intention to purchase cosmetics	0.163.	0.059.	.115	2.776***	.006	
Entertainment → intention to purchase cosmetics	0.194.	0.057.	.170	3.392***	<.001	
Interactivity → intention to purchase cosmetics	-0.045.	0.054.	-.037	-0.847.	.397	62.798*** (.488)
Novelty → intention to purchase cosmetics	0.008.	0.050.	.007	0.152.	.880	
Vitality → intention to purchase cosmetics	0.176.	0.056.	.152	3.134***	.002	
Hypothesis 3 (mediation→dependence)						

Beauty content involvement → intention to purchase cosmetics	0.495.	0.053.	.440	9.295***	<.001
--	--------	--------	------	----------	-------

Note: *p<.05, **p<.01, ***p<.001.

As a result of re-verifying the significance of the mediating effect through the Sobel-test in addition to the verification of the mediating effect of Baron and Kenny previously conducted, the information, entertainment, interactivity, and vitality of the cosmetics purchase are the same as the results of the mediating effect verification of Baron and Kenny. In the influence on intention, the mediating effect of beauty content involvement was verified, as shown in <Table 5> and <Figure 1>.

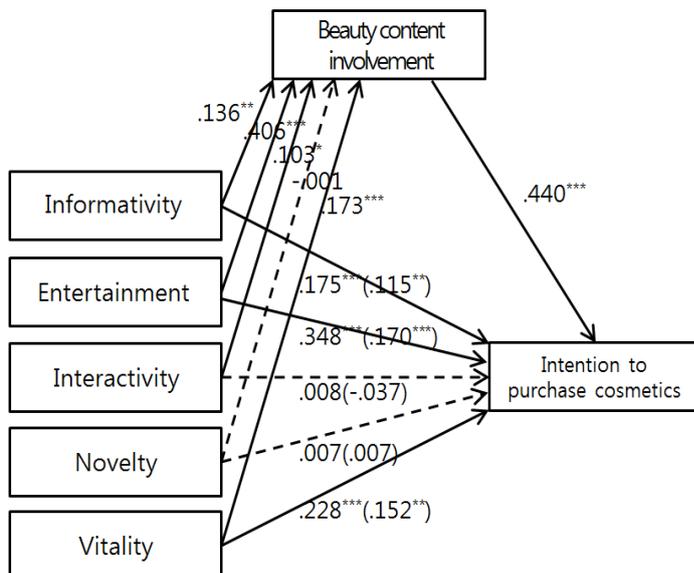
$$Z_{ab} = \frac{a \times b}{\sqrt{(a^2 \times se_b^2) + (b^2 \times se_a^2)}}$$

Table 5. Sobel-test results.

Route	Z	p
Information → beauty content involvement → intention to purchase cosmetics	2.975**	.003
Entertainment → beauty content involvement → intention to purchase cosmetics	6.174***	<.001
Interactivity → beauty content involvement → intention to purchase cosmetics	2.141*	.032
Novelty → beauty content involvement → intention to purchase cosmetics	0.021.	.983
Vitality → beauty content involvement → intention to purchase cosmetics	3.193**	.001

Note: *p<.05, **p<.01, ***p<.001.

Figure 1. The mediating effect of the involvement with beauty content in the relationship between the characteristics of youtube beauty content and the intention to purchase cosmetics.



4.4.2. Verification of the mediating effect of involvement in beauty content for men

For men, a regression model for the relationship between the characteristics of YouTube beauty content and cosmetic purchase intention was appropriate ($F=25.945$, $p<.001$), entertainment ($\beta=.364$, $p<.001$), vitality ($\beta=.333$, $p<.001$), informativity ($\beta=.138$, $p<.05$) in that order. The regression model for the relationship between beauty content involvement and cosmetic purchase intention was appropriate ($F=33.064$, $p<.001$), entertainment ($\beta=.437$, $p<.001$), vitality ($\beta=.240$, $p<.001$), information ($\beta=.131$, $p<.001$) in that order, as shown in <Table 6>.

Table 6. The influence of the characteristics of men's youtube beauty content.

Step	B	SE	β	t	p	F(R ²)
Hypothesis 4 (independence→dependence)						
Informativity → intention to purchase cosmetics	0.817.	0.088.	.138	2.123*	.035	25.945*** (.399)
Entertainment → intention to purchase cosmetics	0.423.	0.078.	.364	5.395***	<.001	
Interactivity → intention to purchase cosmetics	0.065.	0.082.	.054	0.792.	.429	
Novelty → intention to purchase cosmetics	-0.135.	0.084.	-.120	-1.605.	.110	
Vitality → intention to purchase cosmetics	0.394.	0.089.	.333	4.415***	<.001	
Hypothesis 4 (independence→mediation)						
Informationality→ beauty content involvement	0.165.	0.077.	.131	2.136***	.034	33.064*** (.459)
Entertainment→ beauty content involvement	0.470.	0.069.	.437	6.814***	<.001	
Interactivity → beauty content involvement	0.044.	0.072.	.039	0.608.	.544	
Novelty → beauty content involvement	0.000.	0.074.	.000	-0.005.	.996	
Vitality → beauty content involvement	0.264.	0.079.	.240	3.357***	<.001	

Note: * $p<.05$, ** $p<.01$, *** $p<.001$.

Next, as a result of analyzing the mediating effect of beauty content involvement, the regression model was suitable ($F=35.399$, $p<.001$), and the influence of beauty content involvement on cosmetic purchase intention was significant ($\beta=.477$, $p<.001$). The significant influence of entertainment ($\beta=.156$, $p<.001$), vitality ($\beta=.218$, $p<.001$) and information ($\beta=.115$, $p<.001$) was partially adopted as shown on <Table 7>. In other words, in the case of men, among the characteristics of YouTube beauty contents, entertainment and vitality can be said to have an influence on the cosmetic purchase intention by partially mediating the involvement of beauty contents, and the informativity mediates completely the cosmetic purchase intention by affecting the involvement of beauty contents.

Table 7. Results of the verification of the mediating effect of beauty content involvement.

Step	B	SE	β	t	p	F(R ²)
3step (independence → dependence)						
Informativity → intention to purchase cosmetics	0.102.	0.080.	.075	1.277.	.203	35.399*** (.523)
Entertainment →	0.181.	0.078.	.156	2.322*	.021	

intention to purchase cosmetics					
Interactivity →					
intention to purchase cosmetics	0.042.	0.073.	.350	0.578.	.564
Novelty →					
intention to purchase cosmetics	-0.135.	0.075.	-.120	-1.793.	.075
Vitality →					
intention to purchase cosmetics	0.259.	0.082.	.218	3.149**	.002
<hr/>					
3step					
(independence → dependence)					
Beauty content involvement→					
intention to purchase cosmetics	0.515.	0.073.	.477	7.074***	<.001

Note: *p<.05, **p<.01, ***p<.001.

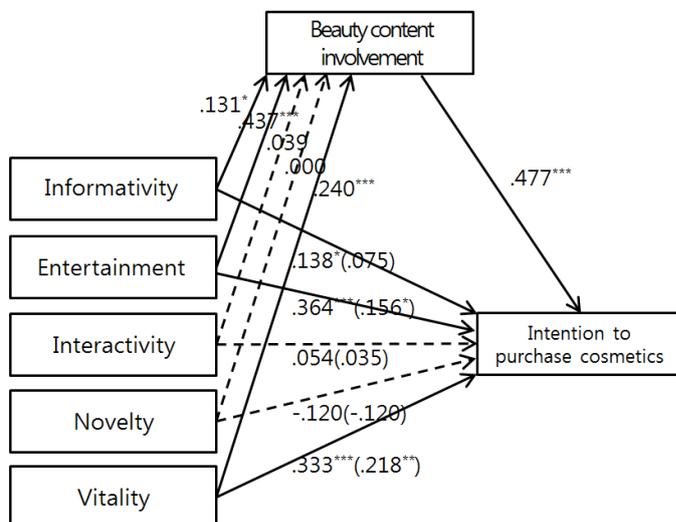
In addition to the previous Baron and Kenny's mediated effect verification, the significance of the mediated effect was re-verified through Sobel-test, and the mediated effect of beauty content involvement was verified in the same way as Baron and Kenny's mediated effect, as shown in <Table 8> and <Figure 2>.

Table 8. Sobel-test results(men).

Route	Z	p
Information → beauty content involvement → intention to purchase cosmetics	2.050*	.040
Entertainment → beauty content involvement → intention to purchase cosmetics	4.900***	<.001
Interactivity → beauty content involvement → intention to purchase cosmetics	0.609.	.543
Novelty → beauty content involvement → intention to purchase cosmetics	0.000.	1.000.
Vitality → beauty content involvement → intention to purchase cosmetics	3.020**	.003

Note: *p<.05, **p<.01, ***p<.001.

Figure 2. The mediating effect of the involvement with beauty content in the relationship between the characteristics of youtube beauty content that uses AI and the intention to purchase cosmetic(men).



4.4.3. Verification of the mediating effect of involvement in beauty content for women

For women, a regression model for the relationship between YouTube content characteristics and cosmetic purchase intention was appropriate ($F=25.387$, $p<.001$), entertainment ($\beta=.269$, $p<.001$), and information ($\beta=.209$, $p<.01$), vitality ($\beta=.175$, $p<.05$), novelty ($\beta=.145$, $p<.05$) in that order. The regression model for the relationship between women's beauty content involvement and cosmetic purchase intention was appropriate ($F=24.726$, $p<.001$), the influence of entertainment ($\beta=.390$, $p<.001$), and interactivity ($\beta=.168$, $p<.05$), informativity ($\beta=.131$, $p<.05$), in that order, was significant, as shown in <Table 9>.

Table 9. The influence of the characteristics of women's youtube content.

Step	B	SE	β	t	p	F(R ²)
Hypothesis 4 (independence → dependence)						
Informativity → intention to purchase cosmetics	0.313.	0.095.	.209	3.278**	.001	25.387*** (.394)
Entertainment → intention to purchase cosmetics	0.301.	0.089.	.269	3.367***	<.001	
Interactivity → intention to purchase cosmetics	0.007.	0.086.	.005	0.078.	.938	
Novelty → intention to purchase cosmetics	0.148.	0.073.	.145	2.028*	.044	
Vitality → intention to purchase cosmetics	0.202.	0.084.	.175	2.402*	.017	
Hypothesis 4 (independence → mediation)						
Informationality → beauty content involvement	0.164.	0.080.	.131	2.045*	.042	24.726*** (.388)
Entertainment → beauty content involvement	0.365.	0.075.	.390	4.857***	<.001	
Interactivity → beauty content involvement	0.177.	0.072.	.168	2.467*	.014	
Novelty → beauty content involvement	-0.001.	0.061.	-.002	-0.023.	.982	
Vitality → beauty content involvement	0.082.	0.071.	.085	1.161.	.247	

Note: * $p<.05$, ** $p<.01$, *** $p<.001$.

Next, as a result of analyzing the mediating effect of beauty content involvement, the regression model was suitable ($F=31.874$, $p<.001$), and the influence of beauty content involvement on cosmetic purchase intention was significant ($\beta=.409$, $p<.001$). The significant influence of entertainment ($\beta=.155$, $p<.001$), vitality ($\beta=.152$, $p<.001$) and information ($\beta=.115$, $p<.001$) was partially adopted as shown on <Table 10>. In other words, in the case of women, among the characteristics of YouTube beauty content, information can be said to affect the intention to purchase cosmetics by partially mediating the degree of involvement in beauty content.

Table 10. Results of the verification of the mediating effect of beauty content involvement.

Step	B	SE	β	t	p	F(R ²)
3step (independence → dependence)						
Informativity → intention to purchase cosmetics	0.233.	0.088.	.155	2.639**	.009	31.874*** (.496)
Entertainment → intention to purchase cosmetics	0.123.	0.086.	.110	1.418.	.158	

Interactivity → intention to purchase cosmetics	-0.080.	0.079.	-.063	-1.007.	.315
Novelty → intention to purchase cosmetics	0.148.	0.067.	.146	2.228*	.027
Vitality → intention to purchase cosmetics	0.162.	0.077.	.140	2.099*	.037
<hr/>					
3step (independence → dependence) Beauty content involvement → intention to purchase cosmetics	0.488.	0.078.	.409	6.273***	<.001

Note: *p<.05, **p<.01, ***p<.001.

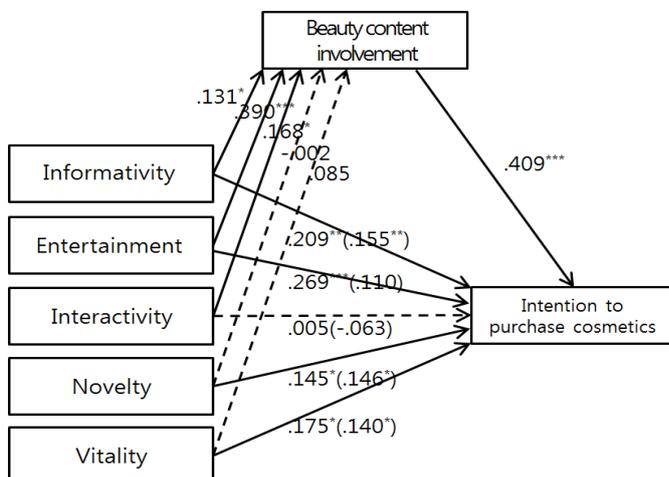
As a result of re-verifying the significance of the mediating effect through the Sobel-test in addition to the verification of the mediating effect of Baron and Kenny previously conducted, the effect of entertainment and interactivity on the purchase intention of cosmetics in the same way as the mediating effect verification result of Baron and Kenny. In the influence, the mediating effect of beauty content involvement was verified, and the mediating effect was verified at the significance level of .01 for informativity, as shown in <Table 11> and <Figure 3>.

Table 11. Sobel-test results(women).

Route	Z	p
Information → beauty content involvement → intention to purchase cosmetics	1.948.	.051
Entertainment → beauty content involvement → intention to purchase cosmetics	3.841***	<.001
Interactivity → beauty content involvement → intention to purchase cosmetics	2.288*	.022
Novelty → beauty content involvement → intention to purchase cosmetics	0.016.	.987
Vitality → beauty content involvement → intention to purchase cosmetics	1.136.	.256

Note: *p<.05, ***p<.001.

Figure 3. The mediating effect of the involvement with beauty content in the relationship between the characteristics of youtube beauty content that uses AI and the intention to purchase cosmetics(women).



5. Conclusion

This study was conducted for the purpose of verifying the factors affecting the cosmetics purchase intention of adult men and women from 20 to 40 years old, who are considered to be the main consumers for marketing effectiveness through YouTube. In order to achieve the research purpose, a self-written questionnaire survey was conducted on 402 adult men and women from 20 to 40 years old nationwide, and the analysis results are as follows.

Among the characteristics of YouTube content, beauty YouTube viewers showed the highest information and lowest interactivity. In other words, it could be said that the motive for using YouTube is to get information. This is consistent with the results of a survey that the purpose of using the Internet in general is to acquire information[11]. In order to secure potential consumers by utilizing beauty YouTubers in companies, it is expected to be effective to produce content that can deliver high-quality information.

As a result of hypothesis verification, hypothesis 1 that says that the characteristics of beauty YouTube content will affect the purchase intention of cosmetics was partially adopted because entertainment, vitality, and information had a significant positive(+) effect. In other words, if the entertainment characteristics are reinforced and interesting images are produced, consumers' intention to purchase cosmetics can be effectively improved. Consumer interest has an important effect on market activation[26]. It is more important to arouse interest rather than producing a boring video focused on information delivery.

The hypothesis 2 that says that the characteristics of YouTube beauty content will affect the degree of involvement in beauty content was partially adopted because entertainment, vitality, informativity, and interactivity had a significant positive effect. The result that entertainment was found to have the greatest influence on the degree of involvement in beauty content suggests an effective direction for producing YouTube beauty content.

Hypothesis 3, that says that YouTube beauty content characteristics mediate beauty content involvement and influence the cosmetic purchase intention was partially adopted as a partial mediating effect of beauty content involvement, was found in the relationship between entertainment, vitality, informativity and cosmetic purchase intention. In particular, the influence of beauty content involvement appears to be very large compared to the characteristics of YouTube beauty content, so it seems more important to increase viewers' involvement in beauty content. In other words, content should be structured to give consumers confidence that they can form a sense of identity or take care of their appearance

Finally, the relationship between the characteristics of YouTube beauty content, beauty content involvement, and cosmetic purchase intentions differed according to gender. Men's entertainment, dynamism, and information mediated their involvement in beauty content to influence their intention to purchase cosmetics, while women's information and entertainment mediated their involvement in beauty content to influence their intention to purchase cosmetics. YouTube algorithm means that YouTube's AI automatically recommends videos in fields that users want to watch or appear to be interested according to big data based on the user's search word. In other words, it is evident that men watch beauty content mainly for interest, and women watch beauty content to obtain the latest information on the cosmetics they will use. Therefore, Youtube's algorithm that uses big data aims to keep viewers in the app for a long time, so in order for viewers to be exposed to the content for a long time and expand their desire to consume cosmetics, interesting content is provided for men, and useful information for women. It suggests what needs to be organized around the content, and in the future, beauty YouTube and big data using information from YouTube can help to develop various information of viewers and advertising strategy in the beauty industry market.

6. References

6.1. Journal articles

- [1] Oh JS. A Study on the Relationship between University Students' Beauty Youtube Viewing Satisfaction and K-beauty. *Journal of Convergence for Information Technology*, 10(12), 175-182 (2020).
- [2] Han SJ. The Effect of Characteristics of Youtube Tourism Contents on Use Satisfaction, Continuous Use Intention and Information Sharing Intention. *Journal of Corporation and Innvation*, 43(3), 155-175 (2020).
- [3] An HK. A Study on the Current Status of Youtube Beauty Channels and Structural Characteristics of Beauty Content Based on Gagne's 9 Instructional Events of Condition of Learning Theory. *Journal of the Korean Society of Cosmetology*, 26(4), 825-841 (2020).
- [5] Lee JW & Choi JS & Yeom JY. Analysis of Properties of Deconstruction & Hair Expression Components in Maison Margiela Collection. *Public Value*, 6(1), 38-50 (2021). [Article]
- [8] Kang YD. A Creative Teaching Method Combining Media Contents Technology with English and American Literature. *International Journal of Human & Disaster*, 4(2), 15-24 (2019). [Article]
- [9] Lee JH & Kang MB & Jeong YD. The Effect of Selection Attributes of Youtube Tourism Channel on User Satisfaction and Intention to Use. *Korean Tourism Industry Research Association*, 45(2), 365-382 (2020).
- [10] Park WS. A Study on Improvement of Information Sharing on Terror Crime. *International Journal of Terrorism & National Security*, 1(2), 1-6 (2016). [Article]
- [12] Shin HB. Developing Sport Legacy through Post-use of Mega-sport Event Facilities: The Case of Winter Sport. *Public Value*, 5(1), 17-28 (2020). [Article]
- [13] Zeng JN & Kim IJ. The Analysis of Affecting Factors on Purchase Intention of Cosmetics on Youtube Beauty Channels. *Korea Internet Electronic Commerce Association*, 18(6), 409-425 (2018).
- [14] Hong LY & Yu SM & Lee JU. The Effect of Brand Crisis Type on Consumer's Perceived Risk and Negative Response in the Context of Online Community: Moderating Role of Information Vividness and Consumer Extroversion. *Korea Advertising Society*, 28(5), 117-144 (2017).
- [15] Lee JH. Interactivity of New-media Contents with Cultural Semiotics Analysis. *Journal of Digital Contents Society*, 21(3), 505-512 (2020).
- [16] Kweon SH. Interactivity Dimension Study in the Internet Media -A Study of User's Recognition Differences of Interactivity Characteristics by Internet Media Genre-. *Korean Association for Broadcasting & Telecommunication Studies*, 21(2), 46-97 (2007).
- [17] Kim HJ & Jo SS. A Study on the Persuasive Effect of Instagram Native Ads Pursuant to Product Involvement and Message Types. *Korea Speech, Media & Communication Association*, 19(1), 79-109 (2020).
- [18] Kim EJ & Gwak TG. An Analysis of Middle-aged Women's Involvement in Fashion and Beauty Depending on Their Participation in Belly Dance. *Journal of the Korean Society of Costume*, 62(4), 195-206 (2012).
- [19] Lee J & Choi EJ. A Study on the Effect of Cosmetic Consumption Desire on Purchase Intention of Customized Cosmetics -Purchasing Motivation as a Mediating Effect-. *Journal of the Korean Society for Wellness*, 11, 43-58 (2020).
- [20] Lee MS & Song TI. A Study on Reliability and Purchase Behavior of Customized Cosmetic -Focus on 20 to 50 Aged Female Living in Gwangju-. *Journal of the Korean Society of Design Culture*, 26(3), 283-293 (2020).
- [21] Ajzen I & Fishbein M. A Bayesian Analysis of Attribution Processes. *Psychological Bulletin*, 82(2), 261 (1975).
- [24] Baron RM & Kenny DA. The Moderator-mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182 (1986).
- [26] Kim TS & Lee JB. Analysis of Perception of Naturopathy Convergence and Utilization Satisfaction Perceived by Skin Beauty Industry Workers. *Protection Convergence*, 5(2), 94-102 (2020). [Article]

6.2. Thesis degree

- [6] Song KS. The Effects of Beauty Direction as Related to Youtuber's Video Reviews and Comments on Consumer's Purchase Intention. Woosong University, Master's Thesis (2019).

- [22] Moon SJ. The Effects of Characteristics of Youtube-based Beauty Contents on Contents Satisfaction and Continuous Use Intention. Sungkyunkwan University, Master's Thesis (2019).
- [23] Hwang HS. The Effect of Youtube Beauty Video Characteristics on Women in Their 10s to 20s Beauty Managing Behavior and Product Purchase Intention. Seokyeong, Master's Thesis (2019).

6.3. Books

- [11] National Information Society Agency. 2020 Internet Usage Survey. National Information Society Agency (2021).
- [25] Kline RB. Principles and Practice of Structural Equation Modeling. Guilford (2005).

6.4. Additional references

- [4] <https://www.igaworksblog.com/> (2020).
- [7] <https://opendict.korean.go.kr/> (2016).

7. Appendix

7.1. Authors contribution

	Initial name	Contribution
Lead Author	HW	<ul style="list-style-type: none"> -Set of concepts <input checked="" type="checkbox"/> -Design <input checked="" type="checkbox"/> -Getting results <input checked="" type="checkbox"/> -Analysis <input checked="" type="checkbox"/> -Make a significant contribution to collection <input checked="" type="checkbox"/> -Final approval of the paper <input checked="" type="checkbox"/> -Corresponding <input checked="" type="checkbox"/>
Corresponding Author*	EK	<ul style="list-style-type: none"> -Play a decisive role in modification <input checked="" type="checkbox"/> -Significant contributions to concepts, designs, practices, analysis and interpretation of data <input checked="" type="checkbox"/> -Participants in Drafting and Revising Papers <input checked="" type="checkbox"/> -Someone who can explain all aspects of the paper <input checked="" type="checkbox"/>

Publisher: J-INSTITUTE
ISSN: 2435-3345

Website: www.j-institute.jp/ai/
Editor: ai@j-institute.jp

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

dx.doi.org/10.22471/ai.2021.6.2.44

© 2021 J-INSTITUTE

Approaches to Forming ETHICAL AI as an Artificial Moral Agent: Suggesting Virtue Education Method Through Comparison of Top-Down and Bottom-Up Approaches

Hyunsoo Kim

Pusan National University, Busan, Republic of Korea

Abstract

Purpose: The purpose of this study is to propose that the virtue education method can be utilized in the process of learning to have ethics as an Artificial Moral Agent (AMA). Recent various discussions about AI presuppose the coexistence of humans and AI. Meanwhile, in order for coexistence to be achieved, it is fundamentally necessary for humans as individual beings to form relationships with others in an appropriate way. This study focused on the use of the virtue education method among the methods for AI to learn the order of relationships as an AMA that becomes the counterpart of human ethical behavior in the interaction between humans and AI.

Method: This study presented basic data using the method of literature research. And based on this, a virtue education method that can be applied to AMA was developed and presented using the method of development research. With this, This research tried to argue about the necessity and validity of the virtue education method.

Results: This study examines that the current discussion on ethical learning in AMA is mainly based on a Top-down approach based on traditional ethical theory. And as an alternative to this, it is suggested that machine learning and deep learning methods can be used as a method for AMA to learn ethics. Based on this, the characteristics and validity of virtue ethics applicable to AMA were presented. And as a method to learn it, a teaching and learning method applicable to the learning of AMA was proposed.

Conclusion: Currently, the approach to preparing ethics for AMA tends to mainly choose a Top-down approach in which a specific ethical theory is selected and the design of a computational algorithm and system that can implement the theory is derived. However, just as children develop moral reasoning through appropriate developmental stages, AMA also needs to actively introduce a bottom-up approach to forming ethical judgment principles. As a specific method, we propose to use the method of grafting the virtue education method of analyzing the characteristics of a virtuous person and repeating it by imitating it to the deep learning method of AI.

[Keywords] Artificial Moral Agents (AMA), AI, Ethics, Deep Learning, Teaching and Learning Method

1. Classification of Ethical Theories and Approaches to AI Ethics

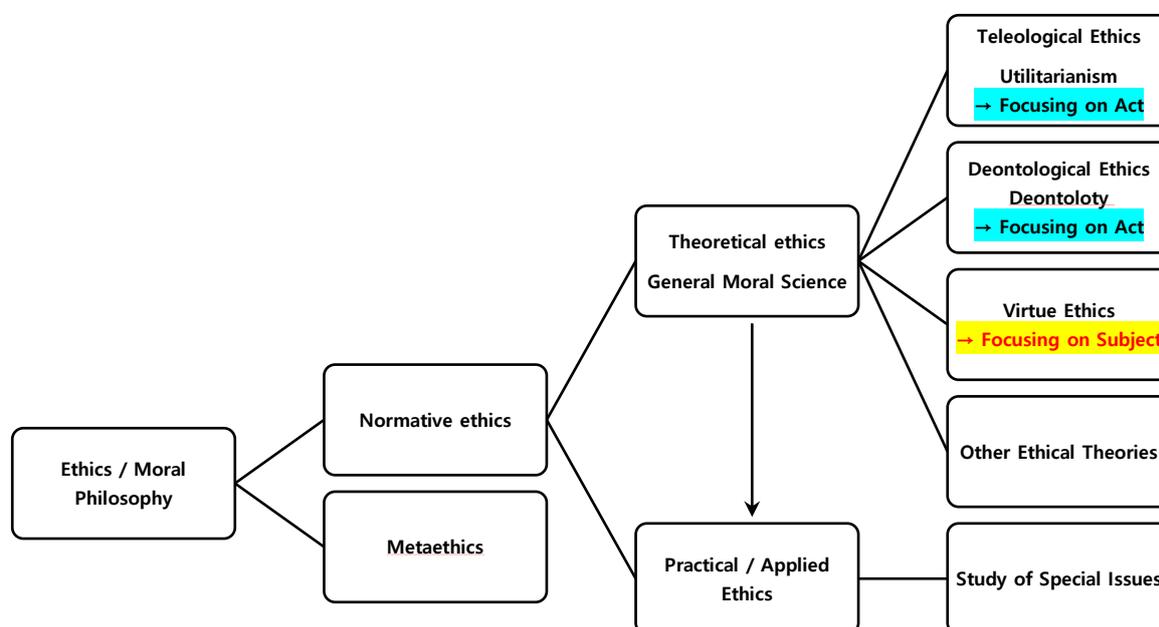
Ethics or moral philosophy is understood as the study of right and wrong. And these ethics are divided into normative ethics and meta-ethics according to their nature. At this time, Normative ethics performs the function of suggesting moral norms that will govern our lives by directly exploring what is good, right, justice, and ought to be. On the other hand, Meta ethics performs the function of analyzing the meaning of moral language and examining the validity of moral reasoning in order to explain the nature of objectivity presumed to be possessed by moral judgment.

Meanwhile, Normative ethics is further subdivided into theoretical ethics and practical ethics. Theoretical ethics aims to establish ethical theories by clarifying the standards or norms of moral judgment, and to guide actions while justifying them. Therefore, Normative ethics try to understand and explore moral norms that act as objective and universal purposes and rules. On

the other hand, Practical ethics aims to solve various ethical problems raised in the realm of modern life based on the principles and contents of Theoretical ethics. Therefore, it is also called applied ethics in that it performs the function of solving problems by applying ethical theory to real moral problem situations.

The representative theories of Theoretical ethics, which is the basis of applied ethics, are utilitarianism that emphasizes results as teleological ethics, deontology as ethics that emphasizes motives, and virtue ethics that focus on diverse virtues. In addition to this, there are various other ethical discussions, including ethics of care, ethics of responsibility, ethics of discourse, and so on. In this case, teleological ethics and deontological ethics have something in common in that they focus on ethical behavior. On the other hand, in the case of virtue ethics, it has a characteristic distinguishing it from other ethical theories in that it focuses on specific habits and behavioral characteristics of actors or subjects with certain positive characteristics. This Analysis could be shown as following <Figure1>.

Figure 1. Classification of ethics.



Currently, various prior studies related to the formation of ethical norms in AI can be largely analyzed as Top-down Approach, Bottom-up Approach, and Hybrid Approach that mixes these two approaches[1]. To show the approach itself made from these two aspects of Top-down and bottom-up, there is a study that attempts to reveal the meaning of human cognition by approaching it analytically[2]. The Top-down approach is a deductive method that suggests ethical principles and applies them to specific cases, and the bottom-up approach is an inductive method that extracts ethical norms through judgments made in individual situations. On the other hand, since the Hybrid Approach itself is presented in a combined form in the process of applying the two approaches rather than having any characteristics. Therefore it would be reasonable that the two approaches of Top-down and Bottom-up can be practically contrasted and reviewed.

Studies that attempt to construct AI models tend to prefer a Top-down approach, and this approach is being used in various fields. There are some representative examples that include a study that proposed Top-down deduction in the primary order of a logical decision-making structure[3], and a research survey that performed Top-down induction in the classification of decision trees[4], and so on.

In the study of AI Ethics, a Top-down approach tends to be preferred. Followings are some examples as a study on AI Ethics related to individual human rights[5], an approach to character

education based on AI[6], a study that explored the ethical challenges of AI required in the era of IR 4.0[7], A study focusing on the issue of Ethical Principles to be implanted in AI-equipped combat robots[8], and so on. Looking at the Top-down approach from the perspective of AI Ethics, it basically has a method of selecting a specific ethical theory and drawing the design of a computational algorithm and system that can implement the theory. The reason why such a Top-down approach is widely used is because it is determined that it can be helpful to solve various, complex and unpredictable real problems based on certain abstract principles and applying them to specific problems.

However, learning in AI actually works in a Bottom-up fashion. As a representative method related to AI learning, Machine Learning is a study on computer algorithms that improve automatically through experience, and is a field that develops algorithms and technologies that enable computers to learn. Specifically, this mainly means developing an algorithm that allows the machine to learn from data and execute operations that are not specified in code. However, Deep Learning, which goes one step further, is used as a term to mean a set of machine learning algorithms that attempt high-level abstraction through a combination of several nonlinear transformation methods. In this case, high level abstractions refer to the work of summarizing core contents or functions in large amounts of data or complex materials. Therefore, such Deep Learning can be understood as a field of machine learning that teaches human thinking to computers.

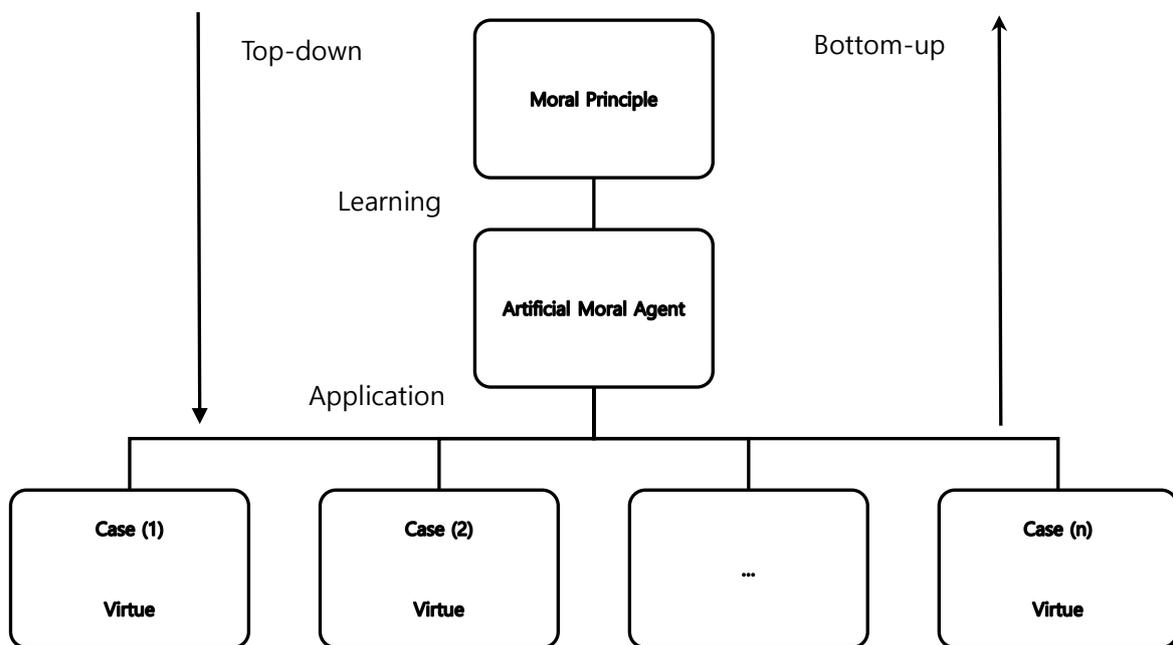
The classification of ethics discussed above can be understood in contrast to the two approaches related to the construction of an ethical model of AI as follows: Firstly, two theories corresponding to action-oriented ethics that provide the standard of right behavior, utilitarianism corresponding to teleological ethics, and Kantism corresponding to deontological ethics, can be seen as the Top-down approach. Utilitarianism is the interpretation of concrete events based on the ethical principle of the greatest happiness of the greatest number, and Kantism is to make moral judgments about specific issues based on the Categorical Imperative that reinterprets the Golden Rule. The Top-down Approach is an attempt to construct an ethical model of AI through activities that devise this type of absolute rule and apply the standard of correct behavior in accordance with it uniformly in all situations.

In contrast with this, theory of virtue ethics corresponds to an agent-centered ethics that analyzes the characteristics of an actor with outstanding virtue and approaches by extracting a specific list of virtues from the characteristics of individual actions that he performs daily, and this can be seen as a Bottom-up approach. This virtue ethics can be understood as an attempt to embrace the richness and diversity of human experience that the formal approach cannot capture. Bottom-up Approach is an activity that extracts positive traits from the figure of a virtuous person and imitates them, or an attempt to construct an ethical model of AI through activities that allow AI to have the characteristics of actors deemed desirable.

The distinction between Top-down approach and Bottom-up approach can be interpreted as a relationship between deductive reasoning and inductive reasoning from a logical point of view. The Top-down approach to AMA can be said to be a kind of Deductive Argument based on Utilitarianism or Deontology. On the other hand, the Bottom-up approach to AMA can be said to be an inductive argument that composes moral principles centered on various virtues that can be grasped and acquired through each specific situation and context and human relationships. These two approaches are shown in <Figure 2>.

This study focuses on the actor-centered virtue ethics in the cognitive process in which AI as an AMA makes ethical judgments, and attempts to demonstrate that it is necessary for the learning process of AI. For this, understanding of virtue ethics and related education methods for the formation of morality in human children are reviewed first. And it would be demonstrated that learning in AMA can be done in a way that imitates the moral development of human children. This will be developed in the form of explaining that it can be achieved through the method of virtue education. These two approaches are shown in <Figure 2>.

Figure 2. Comparison of top-down and bottom-up approaches.



2. Implications of Forming Morality through Virtue Education on AMA

The meaning of the word “Virtue” in Virtue ethics can be said to refer to the internal and external characteristics of an agent with a superior or excellent personal character. Therefore, it can be said that this is an actor-centered ethics that is concerned with being rather than doing. For this, the virtue ethics place more importance on the question of ‘What kind of person should I be?’ rather than ‘What kind of action should I take into?’. The virtue ethic also regards “arête” oriented concepts such as good, excellence, etc., as being of more important basis than duty-oriented concepts such as right, duty, and responsibility.

Virtue ethics reject the idea of being able to codify rules or principles that can provide guidance for particular behavior. This distinguishes Virtue ethics into Utilitarian and Kantian ethics, and is a feature that makes it the third mainstream in classification of ethics. Of course, It is admitted that Utilitarian ethics and Kantian ethics also deal with virtue in a way[9]. And there are scholars who practice the distinction between strong virtue ethics, which gives absolute superiority to virtue itself, and moderate virtue ethics, which does not[10]. In spite of this, in the context of discussion related to moral theory, it is generally understood that Virtue ethics has a certain status, and this serves as the theoretical basis for Virtue education.

In the moral education of children, the approach of Virtue education is functioning as an approach corresponding to the autonomy approach. First, moral education based on the autonomy approach aims to foster rational autonomy based on the reason each individual possesses. By allowing these individuals to make rational moral judgments, education in the form of emphasizing individual autonomy is achieved. However, the approach to virtue education is based on the understanding that the cultivation of morality is the cultivation of meaningful virtue. Therefore, it is intended to provide education focusing on acquiring and developing virtue as an attribute of personality. Virtue in the ethical approach is used to mean communal value of virtues that have traditionally been valued and delivered meaningfully. Therefore, educating these virtues is sometimes understood as a matter of character and behavior.

The approach of Virtue education based on the theory of virtue is embodied in a way to extract specific virtues that have important meaning in society and community, and to learn and achieve the virtues in way of repeatedly habituation of behavior. This approach to Virtue education can contribute to the learning of AMA in that approach can reflect the complex characteristics of human morality. Actually, Top-down approach of setting a specific moral principle first and then solving all situations in which moral problems arise cannot cover all moral problems that occur in reality, and This Virtue education could solve to some extent in the sort of difficult in Top-down approach.

In the Top-down approach, there may be conflicts between the principles it establishes, and the application of mechanical principles may cause problems that conflict with the sense of morality possessed by humans. However, since the approach of Virtue education basically assumes the actor as a personal subject judged to be desirable or excellent, and grasps and approaches the personality characteristics of the actor as a virtue. So it is possible to flexibly cope with the various and complex characteristics of human personality. This is to provide the possibility for AMA to establish a model with desirable personality and to grow into a more human-like subject through learning about the various types of virtue of that personality. If so, it is necessary to examine how these implications can be applied to actual AMA learning. If the learning of AMA is viewed from the perspective of a Bottom-up approach, it will proceed in a form similar to the learning mechanism of children. Just as children expand their cognitive prospects by experiencing individual events, AMA will also be able to develop cognitive mechanisms related to moral judgments through the process of repeatedly learning and practicing judgments made in individual moral situations.

3. Characteristics of AMA and Learning Method of AMA as Human-Mimicking AI

AMA is used as a term to refer to AI that makes its own moral decisions. This means that AI will have advanced autonomous systems that can use moral criteria to evaluate different course of action. It's a question of how to design morally sensitive systems. Studies suggesting that consciousness is necessary for AMA to function properly[11], On the one hand, it is argued that it is impossible to achieve at the current level of technological development due to the complex characteristics required by AMA[12], A study that explores the possibility of the Moral Turing Test for the possibility of AMA[13] has been proposed. There are also studies that point out the risks of AMA in connection with the risks of Artificial General Intelligence(AGI)[14] [15]. Together with this, research that exploring the potential of AMA in relation to the positive aspects of AI[16] and survey on the current status of AMAs[17]also exists.

These studies, while acknowledging that AMA needs to be developed, point out the difficulties it has. However, the characteristic inherent in these difficulties is that it is necessary for AMA to show human-like performance, but it can be said that it is difficult to search for a specific method as to how this should be done.

In order to solve these difficulties, an approach through learning that enables AMA to have human-like ethical judgment ability or function as an ethical actor is required. In terms of moral psychology, this learning can be understood as focusing on learning for the formation of AI that mimics desirable and moral aspects of human beings. The learning of AMA at this time can be designed like the process in which children learn morals, which can be embodied in the form of AMA using deep learning technology to develop its own moral code.

Recent discussions on ethical issues in the future society predict that AI will act as a human representation when it has Autonomous Self-realization[18], and such learning is analyzed to be possible through Deep Learning and Big Data Based AI[19]. And it is understood that the AMA operating as a result of such learning can properly function to solve problems in various realities, including even AI-Based cyber security Issues[20].

Based on the above discussion, the practical advantages that can be obtained when applying the method of virtue education as a Bottom-up method to the moral formation of AMA are as follows. First, since morality can be formed from the initial stage of human development, moral education is possible from the basic level of AI or AMA. This point of view has significance in terms of achieving educational accountability because it is based on the view that ethical education targeting AI can be sufficiently conducted even when AI technology is not advanced to a certain level. This point of view has significance in terms of achieving educational accountability because it is based on the view that ethical education targeting AI can be sufficiently conducted even when AI technology is not advanced to a certain level.

If such a Bottom-up virtue education method is designed in connection with AI deep learning method, it can be proposed as follows.

First, let the AMA extract the desirable virtues required by the AMA itself. At this time, the extraction-related method is first made in the form of Supervised Learning, in which an ethics expert instructs how AMA can extract ethical virtues through everyday examples. This can be called rule-based Knowledge Engineering Learning. However, at this time, there is a point where Supervised Learning is differentiated from the Top-down approach. The Top-down approach is to directly present the moral principle and learn the reasoning method or mechanism of reasoning based on it. On the other hand, Supervised Learning conducted in the bottom-up approach is to suggest what virtues and how to make ethical decisions based on them instead of these mechanisms.

Second, the virtues are extracted through deep learning of AMA. This extracts input through natural language analysis for numerous real and hypothetical cases related to moral issues that exist on the Internet, and analyzes weights by targeting and mapping them in response to the virtues presented through Supervised Learning. Transformation in this layer is configured to express the weight of individual virtues exercised in a specific situation where moral judgment is requested as a function with parameters. At this time, the function is the correction by a human ethics expert with respect to the difference between the prediction result predicted by AMA's Deep Learning and the actual result or the result presented by a human ethics expert for the result of moral action according to the virtue. Intervention should be made to make this happen.

Third, the functions composed of Supervised Learning and Deep Learning are operated as one Policy Network. And based on this, the Value Network is operated for specific situations in which the extracted virtue is applied. Together with this, Reinforced learning that takes place in this process is closely related to AI that learns by accumulating experience. It can be said that this is in line with the virtue education method in moral education in that virtue education is basically more important than practical experience and habituation. In addition, if the Monte Carlo Tree Search method is paralleled as a complement to Deep Learning and Reinforced Learning, it is thought that the ethical learning of AMA will be more valid.

4. Conclusion

The technology of deep learning for AMA can be configured with certain differences from the technology of Machine Learning. Machine Learning means performing a function with given data and gradually improving that function over time. From this point of view, Machine Learning is considered suitable for a Top-down approach. In comparison, Deep Learning is fundamentally understood as a sub-concept of machine learning, but nonetheless it can be said that it is designed to continuously analyze data using logical structures similar to the way humans draw conclusions.

In AMA's education, designing programs to build sharp intelligence and intuition is the key factor. Such Deep Learning is to organize algorithms hierarchically so that they can learn on their own and make intelligent decisions. At this time, various ethical problems encountered in

reality will perform the function of supplying data to Deep Learning algorithms. In that the algorithm of Deep Learning predicts the result through learning based on the data, as the Virtue ethics approaches, it enables the AMA to conduct learning based on the numerous data possessed by the characteristics of people with excellent morality or suitability can be expected.

The design related to the learning of this AMA is also connected to the prediction through the construction of several hidden layers related to morality. At this time, it is judged that the utilization of multilayer neural networks possessing morality and finding the elements of morality related to data analysis will play an important function in the development of AMA in the future. needs to be developed.

5. References

5.1. Journal articles

- [1] Allen C & Smit I & Wallach W. Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches. *Ethics and Information Technology*, 7, 149-155 (2005).
- [2] Dennett D. Cognitive Science as Reverse Engineering Several Meanings of “Top-down” and “Bottom-up”. *Studies in Logic and the Foundations of Mathematics*, 134, 679-689 (1995).
- [3] Blockeel H. Top-down Induction of First Order Logical Decision Trees. *AI Communications*, 12(1-2), 119-120 (1999).
- [4] Rokach L & Maimon O. Top-down Induction of Decision Trees Classifiers -A Survey. *IEEE Transactions on Systems, Man, and Cybernetics*, 35(4), 476-487 (2005).
- [5] Li Y & Park G. AI Ethics and Privacy Right. *Robotics & AI Ethics*, 5(2), 27-33 (2020). [Article]
- [6] Kim Y & Park G. Rubrics and Schoolwide Approach to the Character Education and Some Implications to AI-based Character Education. *Robotics & AI Ethics*, 5(2), 19-26 (2020). [Article]
- [7] Kim T & Park G & Seo E. IR4.0 and Ethical Tasks of AI. *Robotics & AI Ethics*, 4(2), 6-13 (2019). [Article]
- [8] Kim H & Park G. Ethical Issues on AI Equipped Combat Robots. *Robotics & AI Ethics*, 5(2), 1-7 (2020). [Article]
- [9] Nussbaum MC. Virtue Ethics: A Misleading Category?. *The Journal of Ethics*, 3, 163-201 (1999).
- [10] McAleer S. An Aristotelian Account of Virtue Ethics: An Essay in Moral Taxonomy. *Pacific Philosophical Quarterly*, 88, 208-225 (2007).
- [11] Wallach W & Allen C & Franklin S. Consciousness and Ethics: Artificially Conscious Moral Agents. *International Journal of Machine Consciousness*, 3(1), 177-192 (2011).
- [12] Hew PC. Artificial Moral Agents are Infeasible with Foreseeable Technologies. *Ethics and Information Technology*, 16, 197-206 (2014).
- [13] Allen C & Varner G & Zinser J. Prolegomena to Any Future Artificial Moral Agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3), 251-261 (2010).
- [14] Muller VC. Risks of General Artificial Intelligence. *Journal of Experimental & Theoretical Artificial Intelligence*, 26(3), 297-301 (2014).
- [15] Torres P. The Possibility and Risks of Artificial General Intelligence. *Bulletin of the Atomic Scientist*, 75(3), 105-108 (2019).
- [16] Omohundro S. Autonomous Technology and the Greater Human Good. *Journal of Experimental & Theoretical Artificial Intelligence*, 26(3), 303-315 (2014).
- [17] Cervantes JA & Lopez S & Rodriguez LF & Cervantes S & Cervantes F & Ramos F. Artificial Moral Agents: A Survey of the Current Status. *Science and Engineering Ethics*, 26, 501-532 (2020).
- [18] Kim T & Park G. Future Oriented Smart Society and Ethical Issues. *Robotics & AI Ethics*, 6(1), 35-44 (2021). [Article]
- [19] Kim H. Suggestion of Building the AI Code of Ethics through Deep Learning and Big Data Based AI. *Robotics & AI Ethics*, 6(1), 29-34 (2021). [Article]
- [20] Lee S. AI-based Cybersecurity: Benefits and Limitations. *Robotics & AI Ethics*, 6(1), 18-28 (2021). [Article]

6. Appendix

6.1. Authors contribution

	Initial name	Contribution
Author	HK	<ul style="list-style-type: none">-Set of concepts <input checked="" type="checkbox"/>-Design <input checked="" type="checkbox"/>-Getting results <input checked="" type="checkbox"/>-Analysis <input checked="" type="checkbox"/>-Make a significant contribution to collection <input checked="" type="checkbox"/>-Final approval of the paper <input checked="" type="checkbox"/>-Corresponding <input checked="" type="checkbox"/>-Play a decisive role in modification <input checked="" type="checkbox"/>-Significant contributions to concepts, designs, practices, analysis and interpretation of data <input checked="" type="checkbox"/>-Participants in Drafting and Revising Papers <input checked="" type="checkbox"/>-Someone who can explain all aspects of the paper <input checked="" type="checkbox"/>

6.2. Funding agency

This work was supported by Pusan National University Research Grant, 2019.