Preliminary communication / Prethodno priopćenje
Received / Zaprimljen: 30. 9. 2024.
https://doi.org/10.21860/j.15.2.13

Dasom Kim*

Political AI and the General Will of Rousseau[1]

SUMMARY

In this paper, I aim to compare Rousseau's concept of the General Will with generative AI based on artificial neural network deep learning algorithms from the perspective of a rule-based ethical framework. To this end, I focus on Rousseau's issue of the "formation, concentration, and fulfillment of the General Will" to explore the implications of AI use for democracy, particularly in the contexts of democratic decision-making and public policy formulation.

As an alternative for realizing the General Will in lawmaking and public policy development, AI can be considered for gathering public opinion and facilitating decision-making processes. AI-driven opinion-gathering and decision-making can overcome the practical challenges of forming the General Will in democratic systems, including conflicts between majority and minority groups. Furthermore, unlike humans influenced by partisan loyalty or political interests, AI can identify the best policies for everyone in an unbiased manner, fostering broad agreement. Additionally, I critically examine potential issues arising from the politicization of AI, despite its advantages in addressing the weaknesses of democratic systems.

Keywords: General Will, generative AI, political AI, legislation, public policy.

AI AND DEMOCRACY

The impact of AI and robots on humans and society is no longer limited to the technical realm. From the writing of academic papers using big data to addressing complex political and social issues, it is hard to find areas where AI's intervention and influence are absent. Moreover, AI technology is advancing at an astonishing pace.

* *Correspondence Address*: Dasom Kim, The Institute of Chung-Ang Philosophical Studies, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, 06974, Seoul, Republic of Korea. E-mail: bambisom@cau.ac.kr. ORCID: https://orcid.org/0009-0009-2838-7554.

[1] This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023S1A5C2A02095437).

Surveillance and control through AI are increasingly expanding into both personal privacy and public activities. If we do not proactively prevent the problems that may arise from AI's intervention and control, humanity may find itself facing unforeseen risks and crises.

Generative AI, such as ChatGPT, is fundamentally a system of information and rules, as well as a formal system of coding, programming, and deep learning algorithms. This version of Deep Learning 2.0 learns from big data autonomously, based on the structure of artificial neural networks, and processes tasks seamlessly through precise and accurate calculations and rapid judgments. Moreover, with the development of Generative AI, which can be regarded as an integration of human reason and will, the advancement of various types of problem-solving AI is expected to accelerate. The work of interactive AI, such as ChatGPT, closely resembles the decision-making and rational choice processes found in democratic societies.

Generative AI, based on artificial neural networks and deep learning algorithms, solves problems through data collection, analysis, understanding, and interpretation related to the 'presented task'. This process can be compared to the decision-making involved in public policy formulation. Countries that have increased the productivity and efficiency of administrative tasks through the electronic processing of information services and e-business are also striving to transform from 'Electric Government' to 'Intelligent Government'. This transformation is not limited to simply enhancing productivity, efficiency, speed, and accuracy in business processes; it is increasingly recognized as an important means of addressing policy issues and further emphasizes the need to strengthen communication and cooperation with citizens.

There are various models of democracy (Fuchs, 2023, pp. 230–231), but Rousseau, in the 18th century, who advocated for the democracy of the sovereignty of the people, established through his 'general will theory' that legislation and government should embody 'the general will' (fr. *volonté générale*) of the people. The general will that Rousseau refers to is closely related to the public interest. Rousseau's theory of the sovereignty of the people states that all laws cannot be recognized as legitimate unless they represent the general will of the people. Furthermore, the people, as democratic citizens, have the right and responsibility to participate in the formation of the general will. For Rousseau, true freedom means political participation and political autonomy. However, he does not specifically outline an objective procedure or method for confirming whether the laws and policies created through political participation reflect the general will.

Although the principle of majority vote is proposed as a method of democratic decision-making under today's representative democratic system, it has a fundamental limitation in that it cannot fully account for the minority, which occupies the same

proportion as the majority. If the majority is always right or if the minority is not always wrong concerning issues of right and wrong, this implies that the policies supported by the majority may not align with the general will. The same holds true for legislation enacted by legislators. Rousseau's concept of the general will emphasizes the public interest beyond the principle of majority vote, namely the happiness and welfare of the community.

On the one hand, many emphasize that AI is fundamentally different from the human social intelligence required for democratic political discourse and social meaning-making. Some argue that AI is an enemy and a threat to democracy because it is essentially aligned with technocracy (Risse, 2023, p. 64), or can be used to strengthen dictatorship through technological means such as AI (Risse, 2023, p. 57). On the other hand, the potential for AI to support more democratic forms of government and governance is also explored (Coeckelbergh, 2022, p. 66). A thorough evaluation of each of these perspectives should be prioritized; however, it is clear that contemporary representative democracies face numerous challenges. In particular, political scientists point to "citizens' ignorance and the resulting deterioration in the quality of public decision-making," noting that even well-informed voters tend to "choose based on social identities and partisan loyalty" (Risse, 2023, p. 64). In such problematic situations, I believe it is very meaningful to consider whether artificial intelligence can contribute to the operation of democracy (Maeng, 2024, pp. 74–77)

First of all, as an alternative that can reflect the general will in the establishment of legislation and public policy, the use of AI in the processes of collecting public opinions and making decisions can be considered. Opinion collection and decision-making by AI can avoid the practical impossibility of forming the general will and the problem of confrontation between the majority and minorities. In addition, AI can find the best policy for everyone "without prejudice" in a way that everyone can agree upon. To achieve this, it is essential to first confirm or verify whether the data analysis and opinion collection process of AI is fair and whether it accurately reflects the general intentions of citizens. Furthermore, it is crucial to determine whether these intentions genuinely represent the general will.

In this paper, I aim to examine the significance of using AI in the operation of democracy, particularly in relation to democratic decision-making and the establishment of public policy. My focus will be on the issue of "the formation, concentration, and fulfillment of the General Will" that Rousseau proposed. Through this exploration, I will emphasize the importance of preemptive responses to the problems that may arise from the politicization of AI.

DEEP LEARNING ALGORITHMS AND COMMUNICATIVE AI

Recently (September 12, 2024), ChatGPT developer OpenAI released "OpenAI o1", a new version of ChatGPT that enhances its reasoning abilities. ChatGPT is fundamentally a generative AI based on artificial neural networks and deep learning algorithms. It is rapidly evolving into superintelligence as its training data, such as Big Data, becomes more extensive. The current iteration of generative AI, ChatGPT, uses vast datasets to generate new information and generates text by accessing the internet and collecting necessary information. Launched by OpenAI, a company that aims to develop AGI, "OpenAI o1" is an AI model that specializes in solving complex math problems.

Communication of exchanging opinions is an interactive method. The formation and concentration of the general will involves the process of collecting opinions, from which laws or public policies reflecting the general will are formulated. However, in general, the actual legislative process is either unilaterally reflected or compromised, resulting in a compromise. In either case, it is difficult to assert that it reflects the general will. Rousseau defines the general will as a common will, everyone's will, and the public interest. According to this, it seems clear that in order for everyone's will to be reflected, the subjective intentions or objectives of legislation or policymakers must be excluded. To reflect the general will, a fair or neutral third perspective, or an impartial wait-and-see attitude, is required. As the being closest to this point of view at the present time, we can assume an artificial general intelligence (AGI) that enables human-level thinking and can successfully solve problems regardless of the subject. ChatGPT, or its new version, OpenAo1, can be regarded as an AI capable of playing a role comparable to AGI, which is expected to be realized in the near future[2].

The interlocutor for the communicating AI is Big Data, and it can also collaborate with multiple AIs. ChatGPT, a generative AI that uses artificial neural network deep learning algorithms, also works in both directions with tremendous speed of fast computing, unlike conventional computers that only work in one direction. In particular, the artificial neural network structure of rule-based AI works in a way that continues to search from very small rules to higher rules and again to higher rules, just as AlphaGo, who has acquired the principle of Baduk, remembers numerous 'go game records' to find answers and finds new numbers from them if necessary (Powers, 2018, p. 465). Additionally, deep learning is a technology used for clustering or classifying objects or data. Therefore, artificial neural network models

[2] As is well known, AGI can learn and infer like humans and actively cope with problems. Unlike Weak AI, which can only move within learned algorithms, it can cope with new situations that have not been learned. However, since 'human-level thinking' is defined differently by scholars, the specific definition of artificial general intelligence may be different for each field ("What is AGI", 2024).

can be regarded as a system of "similarity-based inference" rules that utilize Big Data (Anderson & Anderson, 2018; p.332; Guarini, 2018, p. 332). These mathematical inference models can be used as ethical inference models.

Given the current trends in AI development, it is highly likely that AI legal assistants will play an active role not only in supporting simple tasks and conducting data research but also in facilitating transaction negotiations, mediating disputes, and making ethical judgments (Kim & Maeng, 2022, pp. 151–152, 156–157). A notable example is the growing movement in certain sectors to implement a judicial system based on strict, mechanical AI for criminal trials, particularly in regions where public trust in the judiciary is low. Furthermore, AI systems capable of self-directed learning through deep learning can act as neutral agents by designing algorithms to provide essential data for policy formulation (e.g., objective data and public opinions). This could enhance the acceptability of policies by ensuring they reflect a consensus that all citizens can agree upon.

In a democratic state, governments responsible for gathering and acting on public opinions must establish and implement policies in line with the will of the general public. However, determining an objective method to confirm which laws or policies genuinely reflect the public will remains a challenge. Communicative AI presents a viable alternative. It can contribute to mitigating or resolving the complex issues democratic societies face in legislative and policy decisions. Traditional democratic decision-making, based on majority rule, often excludes or marginalizes minority interests and can become merely a formal, procedural exercise. It is also prone to political partisanship and bias. In contrast, AI can adopt a third-party perspective, maintaining an impartial and objective stance. Therefore, AI could enhance public trust in government while advancing the democratic ideal of popular sovereignty. By offering unbiased data-driven insights, AI can help eradicate political harms, such as prioritizing special interest groups or creating popular policies solely for political gain. In fact, in order to establish and fulfill democratic legislation and public policy, we can consider the use of more advanced communicative AI (or AGI, possibly in the near future), including ChatGPT or OpenAo1. ChatGPT, as a generative AI, is a drive machine that automatically finds the right words from a sentence structure when an input is provided. Therefore, ChatGPT is an AI that is good at finding the right expression. It specializes in finding the right expression to come next in units of sentences or paragraphs, not in units of words. Since ChatGPT's algorithm is designed to function based on the accuracy of the input, collecting information(knowledge) by using a general search portal aids in obtaining high-quality answers.

As is well known, the current level of AI allows for self-directed deep learning using Big Data based on the structure of artificial neural networks. This AI system

automatically performs precise calculations and makes rapid judgments through this process. When developing new legislation or public policy, Big Data—serving as empirical examples or information—acts as an interlocutor, following predefined rules to generate conclusions that best fit the given input. Once the communication process is complete, AI uses inductive generalization to infer general principles from individual cases. Just as algorithms for language translation can be designed, deep learning algorithms utilizing Big Data function as rule-based ethical reasoning systems. These systems are carefully designed to infer correct answers to complex problems by analyzing individual cases, offering a structured approach to ethical and policy decision-making.

This ethical reasoning machine will infer new public policies that align with the specific ethical situations and conditions provided through input values, utilizing stored Big Data, which includes various ethical cases and historical data. The more accurate and specific the problem situation (policy intention, preferences, goals, costs, available resources, etc.) and input values are, the higher the likelihood of arriving at an appropriate conclusion for establishing new policies. In particular, for new legislation, the system can be expected to reach a conclusion consistent with the legislative intent by incorporating court precedents and rulings, where numerous cases have been accumulated to ensure accurate input values. Additionally, various data with narrative significance, such as historical lessons of success and failure, can serve as a foundation for specific calculations and ethical reasoning (Artz, 2000, pp. 73–78). As such, the reasoning system based on artificial neural network deep learning algorithms can be applied as a framework for ethical reasoning, which is essential for the development of legislation and public policies.

THE GENERAL WILL OF ROUSSEAU AND AI AS A RULE-BASED ETHICAL SYSTEM

Democracy based on the principle of majority rule can never fully embody the General Will. The majority rule as a decision-making method cannot guarantee the rights of minorities and the vulnerable. In fact, the principle of majority rule, which often leads to a monopoly of power or the dominance of the majority's interests, may end up justifying majority rule rather than ensuring fairness. According to Rousseau's theory of the General Will, majority voting is neither a rational nor a just method of decision-making.

In decision-making, the majority principle is a form of inductive reasoning. ChatGPT, a generative AI that combines rule-based learning and machine learning, functions as both a communicative AI and a deductive computational reasoning model. However,

ChatGPT's decision-making process still faces several limitations, such as verbal, gender, and political biases. It is natural to anticipate the development of models that address or resolve these limitations. Such problem-solving models would be rule-based AI systems that provide reasoning to ensure that everyone reaches the same conclusion. In this context, AI, as a system of reasoning rules, can replace subjective communication between multiple individuals with objective communication. In legislative or policy-making contexts, rational communication between political parties or interest groups with specific political agendas is often difficult to achieve. Most legislative processes end with either a superficial compromise between interest groups or a unilateral victory or defeat for one side. These issues highlight the fundamental limitations of a democracy that relies on the principle of majority rule.

As examined earlier, AI as a rule-based ethical system is structurally and systematically very similar to Rousseau's General Will in evaluating the morality of legislation and public policy. Rousseau's General Will represents a moral ideal that all sovereigns must strive to realize. He asserts that the General Will is the will of the sovereign or of all the people. Rousseau's General Will corresponds to a moral ideal that all sovereigns must strive to realize. However, it is difficult to precisely define what the General Will is in practice. Nevertheless, it is equally hard to deny that this concept embodies the most fundamental spirit of democracy. According to Rousseau, "the General Will is always right and tends toward the public advantage; but it does not follow that the deliberations of the people are always equally correct". In this sense, sovereignty is exercised through the General Will, which serves as a measure of whether legislation and public policy truly reflect the people's will. Furthermore, Rousseau's General Will, in this sense, functions similarly to Kant's categorical imperative as a universal principle for evaluating the morality of lawful conduct. It is not merely the sum of the individual wills of those involved in legislation, meaning it is not simply a reflection of the numerical majority.

Rousseau further explains the concept of the General Will by comparing it with other types of wills. He states that "the General Will considers only the common interest, while the will of all takes private interests into account, and is merely the sum of particular wills". Therefore, it is crucial not to confuse the General Will with "the will of all" (fr. *volonté de tous*). In the case of the will of all, it is common for individuals to reject or oppose the demands and orders of the state, but the General Will is a will that cannot be opposed. Moreover, in the case of the majority's will, I may not be part of that majority. In other words, the General Will belongs to a completely different realm from the will of all or particular wills (fr. *volonté particulière*), which are merely empirical facts or unanimous agreements that can be measured numerically. In this sense, the General Will takes on the nature of a moral idea, one that applies to all members of the community without exception, as it is

not an empirical fact that can be observed. As such, the General Will, which cannot be found in the empirical world, is an artificial construct that serves as a standard for measuring the legitimacy of a political community. It represents a political right of a new level, entirely independent of natural law. Therefore, in principle, the legislation and public policies of democratic states and governments must reflect the General Will.

The proposition "The general will is the will of the sovereign, or all the people" defines Rousseau's theory of sovereignty of the people. According to Rousseau, "sovereignty is only an exercise of the general will" and "the general will is always correct and always intended for the public interest, but the will of the people is not always correct." (Rousseau, 1950, p. 26). According to this, the general will is a measure of the morality of the people's will. Therefore, in terms of formality, the general will of Rousseau takes the form of a universal proposition that evaluates the morality of doctors, as Kant did in his categorical order. This general will is not the will of the whole, which means the sum of the special will of the members participating in the legislation, or simply the will of the number majority. However, not one person, on behalf of everyone, has a special will for his or her own interests. "While the general will considers only the common interests, the total will is merely the sum of special wills that consider only the individual interests" (Rousseau, 1950, p. 26). Therefore, it is important not to confuse the general will with the total will.

In the case of the will of all, it is common to reject or oppose the demands and orders of the state, but the general will is a will that cannot be opposed to such. Also, in the case of the majority of wills, I may not be included in the majority. In other words, the general will belongs to a completely different world, unlike the total will, such as the 'particular will', which is merely an empirical fact or unanimous agreement that can be confirmed numerically. In this regard, it can be said that the general will has the character of a moral idea that is valid for all members of the community without exception in that it is an object that cannot be confirmed as an empirical fact. As such, the General Will, which cannot be found in the empirical world, is an artificial association and, is a criterion for assessing the legitimacy of a political community, and relates to an artificially established political right that completely dispels dependence on natural law. Therefore, in principle, the legislation and public policy of democratic countries and governments should reflect the General Will.

According to Rousseau, "the social order is a sacred right which is the basis of all rights. Nevertheless, this right does not come from nature, and must therefore be founded on conventions" (Rousseau, 1950, p. 4). Here, convention is a kind of willful act and a kind of promise or agreement made between humans with freedom of will. Rousseau's social contract based on acts of free will predicts a society without

fraud, deception, and hypocrisy, a society based on mutual trust, in a word, a moral society. In addition, Rousseau's social contract is a moral agreement and commitment that limits individuals' private rights as a kind of alliance to protect everyone's lives and goods, and embodies the General Will, which combines the common will of all members of society. This General Will is 'a moral and collective body' that serves as a legitimate political power and unity (Rousseau, 1950, p. 15), and the state itself as a public person and political body. The state, as such a political unity, can be considered a moral entity with reason and will.

The general will is a universal legislative will that is valid for everyone. The general will is neither an abstract ideal nor a whole quantitative thing. It is like the moral will be held by the people in their capacity as citizens. The general will is, in principle, a reflection of the spirit of democracy. Rousseau argues that "conventions and laws are needed to join rights to duties and refer justice to its object. [...] In the state of society, all rights are fixed by law" (Rousseau, 1950, pp. 34–35), which addresses "the question of legislation". Rousseau repeatedly says, "To make laws are acts of the general will." General will is fundamentally related to "the common good", "the common interest", "the public advantage", "the realization of equal freedom for all", "the rules of justice and peace that all people should follow", and "the will to oppose inequality" in particular (Rousseau, 1950, pp. 23, 26, 30, 36; Shklar, 1985, p. 185). However, it is difficult to confirm which laws or public policy fulfill the general will because the general will is not the total will in number and quantity, but the will on the basis of one's sense of justice.

Rousseau emphasizes that "what makes the will general is less the number of voters than the common interest uniting them," and that "this admirable agreement between interest and justice gives to the common deliberations an equitable character" (Rousseau, 1950, p. 30). Most obviously, in making decisions about legislation and public policy, the general will and the principle of majority seem incompatible. Therefore, interactive AI that can communicate reasonably can compensate for these weaknesses of democracy. ChatGPT can stand in a neutral and uncomfortable position that can gather the opinions of as many people as possible, rather than political or partisan decision-making. AI participating in the legislative process can assume the role of the General Will, while AI lawyers can be seen as legislative agents that indirectly address issues of poverty through their involvement in trial processes and contributions to legal precedents. This possibility heralds the emergence of AI legislators, and its feasibility can be seen as practicing Rousseau's democracy of the formation and concentration of the general will.

BIG DATA AND THE FORMATION AND CONCENTRATION OF THE GENERAL WILL

Democracy is founded on the principle that self-governance is not limited to a select few but is accessible to a large number of individuals. From Rousseau's perspective, democratic citizens must make continuous efforts to achieve the formation and concentration of the General Will. As Rousseau emphasized, political power is not inherently sacred or natural; it originates from the collective will of the people, who are the true sovereigns. This will must be continuously shaped and refined as the will of the state itself. Rousseau argued that citizens are truly free only when they actively participate in forming the General Will, free from the tyranny or coercion of others (Coeckelbergh, 2022, p. 26).

The General Will, as Rousseau refers to it, can be said to be the will to legislate various laws in which the will of all the people is consolidated. The practical implementation of the general will is practically impossible, but AI using Big Data can take over the role. If the big data used to make laws or policies replace the will of the people, AI that performs these tasks, as a rule-based system, can be regarded as a legitimate political power. There have been concerns about the risks of such politicized AI, especially AI combined with big data. For example, there have been long-standing criticisms that it may be exploited for "surveillance capitalism" that facilitates capital accumulation and can empower state totalitarianism (Coeckelbergh, 2022, p. 85; Zuboff, 2015; Zuboff, 2019). In this situation, it is increasingly important to consider the moral use of AI, which is expected to wield greater political power than it does today, as various groups, including international organizations, are advocating.

AI is expected to achieve Artificial General Intelligence (AGI) in the near future. Even at its current level, AI can collect and store Big Data through network technologies such as the Internet of Things (IoT) and cloud computing. The process of analyzing and utilizing this data for problem-solving mirrors the formation and concentration of the General Will through communication, information sharing, and the exchange of opinions among individuals. At this stage, Big Data already demonstrates substantial capabilities in data collection, cognitive processing, and social decision-making. The Big Data generated by IoT and Generative AI, equipped with self-directed deep learning, has evolved to the point where it can independently design new algorithms. This evolution indicates a growing potential for the formation of new opinions as the scope and scale of Big Data inputs continue to expand.

From Rousseau's perspective, the success or failure of democracy depends on active participation in the continuous formation of the General Will. This is because the happiness and welfare of the community represent the social realization of freedom, which requires ongoing effort. Freedom is not something that can be granted or

revoked by others; it must be actively pursued and realized. According to Rousseau, participation in legislative processes to shape the General Will is essential for protecting and achieving this freedom.

Thus, political participation is a fundamental duty for any citizen striving for freedom in a democratic society. Merely adhering to the formal procedural justification of majority rule, while neglecting the process of forming and concentrating the General Will, contradicts the essence of popular sovereignty. However, the challenge lies in the lack of an objective method to ensure that the formation and concentration of the General Will are transparent and fair. If AI can address these critical weaknesses, it could offer an innovative solution for advancing democracy by enhancing fairness, inclusivity, and trust in the decision-making process.

With its deep learning capabilities, Generative AI has the potential to evaluate existing policies and develop new ones that reflect citizens' desires based on Big Data. For example, a U.S. company has developed an Artificial General Intelligence (AGI) called ROBAMA, which is designed as a vital support tool for social and political activities. ROBAMA can survey over 5,000 public opinions daily, functioning as a decision-making support system with the potential to supplement or even replace traditional government and legislative processes. ROBAMA ensures that citizens can freely express their opinions and receive immediate feedback, enhancing public engagement. It also plays a critical role in decision-making by identifying key information on significant social and political issues, analyzing evidence, and generating comprehensive reports. The development of AI systems with such capabilities could facilitate the formation and concentration of the General Will, while also mitigating political biases that often influence legislation, media, and civic organizations (Rousseau, 1979, p. 37). This technological advancement could become a crucial countermeasure to address the ongoing crisis facing democracy, promoting fairness, transparency, and inclusivity in governance.

However, it is crucial to take preventive measures against the potential consequences of decision-making and influence through algorithms. One of the most significant concerns regarding the formation and concentration of the General Will, as well as the use of AI in decision-making, is the handling of Big Data. The misuse or abuse of AI, driven by biases inherent in Big Data and algorithms, can pose a threat to the very fabric of society. Indeed, concerns about these adverse effects have led to a growing demand for "explainable AI" that can minimize or verify misjudgments and errors, particularly regarding the structural characteristics, technical complexity, and reliability of Generative AI. Additionally, it is essential to establish an organization dedicated to supervising data management to prevent the government or a small number of companies from monopolizing or controlling data. Considering the

relationship between AI, the General Will, and Big Data, the future success of democracy hinges on the "moral use of Big Data" and the "development of Explainable AI" that can effectively replace the General Will.

POLITICAL AI AND RELIABLE AI

Democracy involves the process by which public opinions are exchanged, and decisions are made regarding which laws, policies, and institutions should be established. Rousseau states the following:

> How are the people to regulate the conditions of the society? Is it to be by common agreement, by a sudden inspiration? Has the body politic an organ to declare its will? Who can give it the foresight to formulate and announce its acts in advance? Or how is it to announce them in the hour of need? How can a blind multitude, which often does not know what it wills, because it rarely knows what is good for it, carry out for itself so great and difficult an enterprise as a system of legislation? (Rousseau, 1950, pp. 36–37).

As a qualification for such legislators, which can be seen as aligning with the principle of separation of powers, Rousseau emphasizes that, as individuals positioned outside the body of the state, legislators should not use the law as a means to satisfy personal desires (Rousseau, 1950, pp. 37–39). In summary, Rousseau argues that legislators should resemble gods who provide laws for humanity (Rousseau, 1950, p. 38). Paradoxically, Rousseau's demand for legislators to be God-like is literally just a wish. Using Big Data to replace the formation and concentration of the General Will does not solve all problems. Dependence on political AI comes with significant risks. The greater the dependence on Big Data and algorithms, the more serious the civil liberties may be compromised by countries or governments with certain powers that strengthen their control over citizens. The possibility of 'control by AI or algorithmic governance' may pose the most powerful threat to the basic principle of democracy.

In general, the prevailing idea is that excessive dependence on AI weakens or reduces human autonomy and self-determination, which are the basic rights of democracy. This problem can hardly be said to be limited to dependence on AI. This is because, in many cases, it is common to seek help from someone to make one's own decisions. Just as it is undesirable to rely on help from others unconditionally, the greater the dependence on AI, the higher the likelihood of facing unexpected difficulties. In this regard, a good example is the case in which trillions of dollars disappeared when programmatic sales of stocks triggered a massive sell-off within 15 minutes during the 2018 global stock market crash. It was found that 'Algorithmic Trading by AI' was the cause ("Algorithmic trading", 2024). This was an incident that occurred as a

result of AI comprehensively analyzing data and automatically trading stocks when various stock price variables were made into Big Data and entered. Such an event can occur even if it is not necessarily caused by AI. The more serious problem that excessive dependence on AI could cause is its intervention in decision-making related to the lives of the majority of people in ways that cannot be controlled.

Due to biases in Big Data, distorted analyses by AI, manipulation of information through algorithms, and the mass production of false information, individuals' legitimate choices or suffrage may be compromised. This is closely related to the specific policy decisions informed by the data. One can consider cases where irrelevant data are used in policy decisions or where past or favorable data are deliberately selected by policymakers to achieve desired outcomes. These instances are also quite common.

The "moral use of Big Data" is closely related to the monopoly and the potential for misuse or abuse of Big Data. As the concentration and monopoly of data accelerate and AI's data processing capabilities improve, concerns about information dictatorship and surveillance societies have become pervasive. In particular, the 'moral use of Big Data' is directly connected to the issue of reliable AI. Measures must be taken to prevent policy decisions made using Big Data from being planned in advance to benefit a specific group. This is the most undemocratic act of abusing Big Data, corresponding not only to the manipulation of the General Will but also posing a significant risk to democracy itself. Additionally, the monopoly of algorithms and Big Data by powerful groups may facilitate the manipulation of public opinion through propaganda. Thus, guidelines are necessary to enforce the moral use of Big Data and the algorithms that utilize it. This raises concerns about trust in AI; in an increasingly AI-driven society, distrust in AI can contribute to significant social anxiety. Addressing the social problems caused by this will require enormous social costs.

Reliable AI problems can also be addressed through the development of "Explainable AI" (Kim, 2023, pp. 276-277, 286-287). The misuse and bias of Big Data can be partially mitigated through the regulation of algorithms used to tackle these issues. For example, ChatGPT, a generative AI based on an artificial neural network deep learning algorithm, addresses this concern through self-directed learning via its feedback system. However, even in this case, the problem remains unresolved. Most people do not understand how AI makes decisions and cannot verify the fairness of its processes. Considering that AI is not perfect and, moreover, is susceptible to human intentional intervention—in other words, acknowledging the possibility of error in any situation—ignorance of AI's working processes could lead to unforeseeable or irreversible risks. The normal operation of AI's autonomous deep learning is difficult

to understand unless one is a high-level expert. It is impossible to leave such a machine to determine the lives of most people.

To prevent this risk, there is a growing demand for a new concept of AI, or "explainable AI", that can transparently show the process behind AI's judgments and actions or explain them to humans. "Explainable AI" is an AI that can articulate its actions and judgments in a manner comprehensible to humans, explaining how decisions are made (Kamath & Liu, 2021, ix). One measure to enhance the reliability of AI applications is to impose an "obligation to provide explanations of the operation of algorithmic models and their conclusions" on interpretable machine learning and explainable AI. This procedure is necessary for AI to diagnose and debug potential biases in algorithmic models while also clarifying the conditions and calculation processes that lead to their conclusions, thereby enhancing trust in the results.

CONCLUSION

Numerous studies have examined the relationship between AI and democracy. Even before the advent of AI, the so-called crises of democracy—including regression and contraction—were widely discussed. In the context of practicing the General Will, many criticisms have been raised regarding whether public decision-making processes, such as elections and voting, truly reflect the political equality of all citizens. Above all, while democratic elections can facilitate the transfer of power, merely winning an election often fails to foster the formation and concentration of the General Will and may instead intensify political polarization.

In this context, introducing AI into a democratic political system—given its potential for both positive and negative impacts—can be seen as a bold and risky endeavor. However, if the use of AI becomes inevitable not only in politics but in almost all areas of society, our task is to minimize the threats AI poses to democracy while promoting its development and use to enhance democratic processes.

AI has already been employed in fields such as law and politics, including legislative activities and public opinion polling. However, AI has become highly politicized and is particularly vulnerable to further politicization in these domains. Since human actions are inherently influenced by politics, AI systems designed and used by humans are, by extension, also political. Consequently, the political misuse of AI could have severe consequences, as few areas of human life remain unaffected by politics.

Decision-making based on the principle of majority rule often initiates new political disputes and does not necessarily constitute a fair or just political act. Given the inherent limitations of democracy—such as its frequent difficulty in aligning with the

realization of a collective will like Rousseau's concept of the General Will—the ethical use of AI, including the development of explainable algorithms and moral guidelines, will play a critical role in determining whether democracy can succeed alongside AI.

As AI technologies advance in applying and executing reasoning rules, it is crucial to ensure that humans do not become passive subjects of algorithmic governance. To achieve this, explainable AI must be implemented as a system capable of validating the reasoning process to minimize biases and informational errors in Big Data. Additionally, it should enhance the transparency and fairness of decision-making by clarifying how Big Data is collected, stored, analyzed, and processed.

References

"Algorithmic trading". (2024). *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Algorithmic_trading.

Anderson, M. & Anderson, S. L. (eds.). (2118). *Machine Ethics.* Cambridge University Press.

Artz, J. M. (2000). Narrative versus Logical Reasoning in Computer Ethics, in: R. M. Baird, R. Ramsower, and S. E. Rosenbaum, (eds.), *Cyberethics*, Prometheus Book.

Coeckelbergh, M. (2022). *The Political Philosophy of AI: An Introduction*, Cambridge: Polity Press.

Fuchs, C. (2023). *Digital Democracy and the Digital Public Sphere*. New York: Routledge.

Guarini, M. (2018). Computational Neural Modeling and the Philosophy of Ethics. Reflections on the Particularism-Generalism Debate. In: M. Anderson and S. L. Anderson, (Eds), *Machine Ethics.* Cambridge University Press.

Kim, D. & Maeng, J. (2022). "Artificial Consciousness and AI-Robot Artist", *Philosophical Investigation, 67*, Journal of Chung-Ang Philosophical Studies, Korea: Chung-Ang University.

Kim, H. (2023). "Unexplainable Explainable AI", *Synthesis Philosophica*, *76*(2), 275– 295.

Kamath, U. & Liu, J. (2021). *Explainable Artificial Intelligence*: *An Introduction to Interpretable Machine Learning*. Switzerland: Springer.

Maeng, J. (2024). Communicative Power and Democracy of AI, *Journal of Philosophical Investigation*, *74*.

Powers, T. M. (2018). Prospects for a Kantian Machine. In: M. Anderson and S. L. Anderson, (Eds), *Machine Ethics*. Cambridge University Press.

Risse, M. (2023). *Political theory of the digital age: where artificial intelligence might take us*. Cambridge: Cambridge University Press.

Rousseau, J.-J. (1950). *The Social Contract or Principles of Political Right*, in *The Social Contract and Discourses*, translated with an Introduction by G. D. H. Cole, New York: E. P. Dutton and Company.

Rousseau, J.-J. (1979). *Emile or On Education*, Introduction, Translated and Notes by Allan Bloom. Penguin Book.

Shklar, J. N. (1985). *Men & Citizens: A Study of Rousseau's Social Theory*. Cambridge: Cambridge University Press.

What is AGI (Artificial General Intelligence)? (2024). Amazon Web Services. Retrieved from https://aws.amazon.com/what-is/artificial-general-intelligence/?nc1=h_ls

Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* London: Profile Books.

Politička umjetna inteligencija i Rousseauova teorija opće volje

SAŽETAK

U radu namjeravam usporediti Rousseauovu teoriju opće volje s generativnom umjetnom inteligencijom koja se temelji na algoritmima dubokog učenja umjetne neuronske mreže sa stajališta etičkog okvira temeljenog na pravilima.

U tu svrhu usredotočit ću se na Rousseauovo pitanje „formiranja, koncentracije i ispunjenja opće volje" kako bih istražio implikacije uporabe umjetne inteligencije (UI) u svrhu demokracije, osobito u kontekstu demokratskog odlučivanja i formuliranja javne politike. Kao alternativa za realizaciju Opće volje u izradi zakona i razvoju javne politike UI se može uzeti u obzir za prikupljanje javnog mišljenja i olakšavanje procesa donošenja odluka. Prikupljanje mišljenja i donošenje odluka uz pomoć umjetne inteligencije može prevladati praktične izazove oblikovanja opće volje u demokratskim sustavima, uključujući sukobe među većinskim i manjinskim skupinama. Nadalje, za razliku od ljudi na koje utječu stranačka lojalnost ili politički interesi, UI može identificirati najbolje politike za svakoga na nepristran način, potičući široku suglasnost. Osim toga, kritički ispitujem potencijalne probleme koji proizlaze iz politizacije UI-ja, unatoč njegovim prednostima u rješavanju slabosti demokratskih sustava.

Ključne riječi: opća volja, generativni UI, politički UI, zakonodavstvo, javna politika.