

[ADRN Issue Briefing]

## Harnessing AI to Strengthen South Korean Democracy

Sunghack Lim (University of Seoul)

### Background: AI and the Crisis of Democracy in the Digital Age

The field of Artificial Intelligence (AI) has evolved from a mere computational tool to a pervasive force that is profoundly impacting human daily life, public discourse, and political decision-making. The advent of generative AI and the accelerated transition toward agentic AI signify a pivotal moment for contemporary governance. For the Republic of Korea, a nation that boasts world-class digital infrastructure and a vibrant civil society, the dissemination of AI technologies has now surpassed the point of no return. The 2025 Seoul Survey of Seoul metropolitan residents capture a striking duality in public attitudes toward AI (Seoul Metropolitan Government 2025). A survey of the public's perception of potential threats to social safety revealed that deepfakes, AI-generated fake news, and personal data leaks were regarded as the most severe threats, while at the same time, there was an overwhelming expression of support for the implementation of AI-driven public services, including welfare blind-spot detection and customized policy platforms.

Discussions within academic and policy circles concerning the nexus of artificial intelligence (AI) and democracy have historically been characterized by a techno-pessimistic, regulation-centric paradigm, with a predominant focus on the potential hazards associated with surveillance capitalism, algorithmic bias, and electoral manipulation. While these concerns are valid, an exclusive focus on risk mitigation is no longer sufficient. In order to ensure the continued vitality of democratic institutions, it is imperative to undergo a paradigm shift, moving away from a state of unconditional pessimism and toward the active institutional design that will harness the democratizing potential of artificial intelligence. This shift must be accompanied by the development of robust safeguards to mitigate the potential threats to democratic institutions posed by artificial intelligence.

In order to analyze the impact of AI systematically, this briefing distinguishes between two layers of democratic governance. Procedural democracy is predicated on the institutional rules that govern the acquisition and transfer of power. Such rules encompass elections, the rule of law, and administrative fairness. Substantive democracy shifts the focus from procedural rules to outcomes, raising questions about whether citizens possess the conditions necessary for meaningful participation, autonomous judgment, ongoing deliberation, and socioeconomic equality. In the following discussion, the author will draw on the work of Jungherr (2023) to map these elements onto three analytical levels: the individual level (self-rule), the group level (equality), and the institutional level (elections). It

should be noted that the system level (competition between political systems) has been left aside as it lies beyond the scope of analysis.

### **Threats to Electoral Integrity: Procedural Democracy Under Siege**

The cornerstone of procedural democracy is the electoral system, which facilitates the acquisition and transfer of political power through a fair and transparent process. Generative AI poses a direct and escalating threat to this foundation. The advent of deepfake technology has ushered in a new era of automated, mass-scale production of hyper-realistic fabricated audio and video, with the potential to distort voters' perceptions, damage candidates' reputations, and erode public trust in electoral outcomes (Hong 2024). The magnitude of this menace is not merely speculative; it is a tangible reality. During the recent Jeonnam-Gwangju integrated regional primary elections, South Korean authorities unearthed 1,600 instances of illicit electioneering employing AI-generated deepfakes, thereby demonstrating the potential for generative AI to be utilized as a tool to expedite a cognitive crisis of truth (Hankookilbo 2026).

A second threat to the integrity of the electoral system emerges through a different, albeit equally corrosive, channel: political finance. The electoral system's democratic legitimacy is predicated on the premise that all citizens participate on roughly equal terms and that the process of acquiring and transferring political power is governed by votes, not by wealth. The integrity of the electoral process is fundamentally undermined when substantial, untraceable capital from the technology industry flows into electoral campaigns. The result is a drift toward plutocracy, a condition in which financial resources rather than the aggregated preferences of citizens determine who wins office and whose interests govern (Jackson and Woolley 2025). In South Korea, the legal prohibition on corporate and organizational political donations is more stringent than in countries such as the United States, thereby mitigating the associated risk. However, it should be noted that the institutional firewall is not impenetrable. Corporations circumvent the law through "donation splitting" (jjogaegi huwon), a strategy that involves the distribution of funds across a network of small individual contributions, thereby evading detection.

### **Policy Recommendations: Safeguarding Procedural Democracy**

The primary policy recommendation is the establishment of a "Public-Interest AI Defense Infrastructure." Recent advancements in technology have enabled South Korea to make a significant stride in the realm of electoral integrity. In anticipation of the recent local elections, the Ministry of the Interior and Safety (MOIS) and the National Forensic Service (NFS) have initiated the deployment of an advanced "AI Deepfake Detection Model." This innovative model integrates global flow analysis and local artifact detection, achieving a remarkable 97% accuracy rate, as reported by Yonhap News in 2026. Electoral management bodies should proactively adopt analogous technologies and mandate the implementation of digital watermarking and content provenance standards for all AI-generated political content. The National Election Commission of South Korea has instituted a comprehensive prohibition on the utilization of artificial intelligence (AI)-generated deepfake content for campaign purposes during the 90-day period preceding election day. This prohibition encompasses synthetic audio, images, and video indistinguishable from reality (National Election Commission 2025).

The second recommendation pertains to the regulation of political finance. According to International IDEA (2026), electoral bodies around the world are beginning to adopt artificial intelligence (AI) auditing. For instance, the UK Electoral Commission is exploring optical character recognition (OCR) and natural language processing (NLP) to scan financial invoices and identify instances of non-compliance, while Mexico's National Electoral Institute has implemented near real-time expense monitoring. It is imperative that the South Korean National Election Commission (NEC) prioritize the implementation of these initiatives, particularly in light of South Korea's vulnerability to digital and algorithmically assisted election campaign content disseminated on YouTube. The oversight of hidden financial patterns through algorithmic analysis has the potential to mitigate indirect corporate influence and revitalize public confidence in electoral integrity.

The third recommendation calls for the implementation of a dedicated "Electoral Integrity AI Model" to address the broader challenge of securing the electoral process as a whole. In the contemporary electoral context, there is an escalating prevalence of coordinated, cross-border threats that range from domestic fraud to disinformation orchestrated by foreign entities. These threats have the potential to render piecemeal defenses inadequate (Hong 2024). The model's functionality would be characterized by its ability to continuously monitor and analyze abnormal data patterns in real time. This capability would position the model as a comprehensive early-warning system, signifying a commitment to democratic resilience that extends beyond a mere technical upgrade.

## **Substantive Democracy: The Individual and Group Levels**

### **1. Current Status and Risks: Self-Rule and Equality**

While the establishment of procedural mechanisms is imperative, it is insufficient on its own. Substantive democracy posits the question of whether citizens genuinely possess the capacity to govern themselves and participate on equal terms. The potential dangers posed by artificial intelligence (AI) manifest in two distinct yet interconnected ways. On the individual level, AI has the capacity to diminish epistemic agency and autonomous judgment. On the group level, it serves to exacerbate existing structural inequalities in participation and representation.

#### **1.1 Individual Level: Erosion of Self-Rule**

Substantive democracy is predicated on the principle of "self-rule," which posits that citizens possess the requisite information and independent judgment to make autonomous political decisions. The potential threat of artificial intelligence (AI) to self-rule is not through overt coercion, but rather through the more insidious erosion of the epistemic conditions that facilitate genuine self-rule.

The initial mechanism pertains to the erosion of individual informational autonomy through algorithmic shaping. The advent of artificial intelligence has led to an unprecedented deluge of information, yet the design of recommendation algorithms is not oriented towards the promotion of truth. Instead, these algorithms are engineered to capture attention. The result is that algorithms direct individuals toward content with which they already agree, thereby creating filter bubbles that insulate citizens from disconfirming evidence. Moreover, these algorithms surface provocative content from political opponents to maximize engagement, thus reinforcing echo chambers. The crux of the issue

lies not in the reception of biased information by citizens, a phenomenon that has persisted since the dawn of politics. The operation of algorithmic curation is characterized by its invisibility and extensive nature, leading to a gradual erosion of the shared factual reality that forms the foundation of democratic deliberation.

The second mechanism is the gradual drift toward technocracy. Democracy is predicated on a deliberate act of trust, namely the belief that the distributed judgment of ordinary citizens is more legitimate than the concentrated authority of any expert. This phenomenon is particularly salient in the context of AI, where the inherent trust placed in the system can be a source of pressure. As AI systems exhibit remarkable aptitude in anticipating intricate trends, a substantial epistemological transformation is precipitated. The underlying question of whether to allow ordinary citizens to determine the timing of AI-driven optimization processes appears to be a valid one. Technocracy, however, does not emerge through coercion; rather, it is characterized by a seductive logic of efficiency. When politics is conceptualized as an optimization problem, civic deliberation appears irrational, and democratic self-rule gradually erodes as power shifts from the electorate to an unelected technocratic elite (König 2023).

## 1.2 Group Level: Deepening Inequality

At the group level, the central value is substantive equality, which entails more than just the principle of "one person, one vote." It also involves equal access to political voice, representation, and public resources. The structural consequences of AI in this context are of particular significance, yet they are often unseen, as they serve to exacerbate existing social divisions rather than to generate new ones.

The initial two mechanisms through which AI exacerbates inequality are rooted in a shared structural problem and must therefore be understood in conjunction. Machine learning models are trained on historical datasets, namely records of societies that formally and systematically engaged in discrimination. This bias manifests in a paradoxical pattern of the visibility imbalance, in which the same marginalized group can be simultaneously too absent from some datasets and too present in others. The underrepresentation of demographic groups in training data for public services is a significant issue that can lead to systemic exclusion from welfare, employment, and financial access. The algorithmic model is unable to recognize and serve these groups, resulting in systemic bias and marginalization. This invisibility can result in systemic exclusion from AI-driven public services or biased treatment in automated hiring processes. Conversely, historically marginalized groups are often overrepresented in certain datasets, such as crime records. This overrepresentation signifies that they are disproportionately subjected to the adverse consequences of AI-assisted predictive policing or sentencing. Moreover, they are susceptible to biased electoral redistricting, which can have a significant impact on their political representation and influence.

The third mechanism operates through the economy. In contrast to the preceding two, this particular concern pertains to the manner in which artificial intelligence (AI) influences the material conditions of democratic participation. As corporations increasingly adopt AI-driven automation, this has led to the displacement of workers, a erosion of bargaining power among the workforce, and a contraction in income, thereby undermining the socioeconomic foundations that are crucial for the sustenance of substantive democracy. The current wave is characterized by its extensive reach, which distinguishes it from previous automation trends. While automation has historically impacted manual

labor, generative AI now poses a threat to white-collar and knowledge-based professions as well. This phenomenon was exemplified during the 2023 Hollywood writers' strike, which highlighted the potential for generative AI to disrupt not only the workforce but also the creative sector. The concentration of artificial intelligence (AI)-generated prosperity among a select group of technological experts has the potential to exacerbate economic and political inequality by creating a mutually reinforcing cycle. Those who possess the necessary resources to influence the development of artificial intelligence (AI) tend to gain disproportionate political influence. In contrast, those who are displaced not only experience a loss of income but also the civic standing that economic security can facilitate.

## 2. Overcoming Risks and Harnessing AI for Substantive Democracy

If left unaddressed, the current trajectory of AI will continue to reshape democratic societies in ways that concentrate informational power, narrow the scope of political judgment, and entrench existing inequalities. However, this trajectory is not inevitable. The technological capacity that renders AI a potential threat to substantive democracy can, under different institutional conditions, become a resource for strengthening democracy. In this case, AI would transition from an instrument of cognitive manipulation, technocratic domination, and structural inequality to a democratic infrastructure for empowerment, inclusion, and augmented deliberation.

### 2.1 Restoring Individual Self-Rule

The initial step in this inquiry is to identify the conditions under which the distortion of epistemic agency by algorithmic processes goes unchecked. The recommendation algorithms that influence the content consumed by millions of citizens are opaque, operating in a manner that is not subject to public scrutiny or democratic oversight. To address these concerns, governments must enforce algorithmic transparency requirements and mandate regular, independent external audits of major digital platforms. However, transparency alone is not sufficient for fostering genuine deliberation; it merely identifies issues without establishing the necessary conditions for constructive dialogue. In this context, Civic Tech emerges as a pivotal entity. Rather than utilizing artificial intelligence (AI) to manipulate preferences, governments can repurpose it as an instrument of augmented deliberation. A particularly illuminating example is provided by Taiwan's vTaiwan initiative, which utilizes the Pol.is algorithm. This initiative maps the landscape of public argument, clusters myriad positions, identifies cross-partisan agreement, and surfaces "rough consensus" while marginalizing toxic exchanges (Yang 2026). In contrast to the maximalist approach of commercial algorithms, which prioritize emotional engagement, vTaiwan's architecture fosters reasoned exchange, thereby rebuilding the epistemic conditions necessary for self-rule.

The second recommendation is intended to address the potential threat of technocracy. It is imperative that AI is not conceptualized as an autonomous oracle that adjudicates political questions; rather, it should function as a mediating infrastructure that serves to inform, rather than supplant, democratic deliberation. In order to establish a formalized framework for this principle, governments must institute "Digital Citizens' Assemblies," which are deliberative bodies entrusted with the responsibility of evaluating AI-driven policy recommendations and supervising the implementation of high-impact AI systems. This arrangement incorporates a human-in-the-loop mechanism, thereby

ensuring that a human decision-maker remains in the chain of authority before any AI-generated recommendation becomes binding policy. This mechanism serves to preserve popular sovereignty against the quiet encroachment of technocratic governance.

## 2.2 Ensuring Group-Level Equality

As artificial intelligence (AI) perpetuates structural inequality through accumulated architectural bias rather than any single act of discrimination, the first and second group-level policy responses must address both the timing and the direction of intervention simultaneously. With regard to the matter of timing, governments must transition from a post-hoc detection model to an ex-ante human rights audit model. These assessments are meticulously conducted prior to the deployment of any AI system in public services. They are designed to identify and rectify discriminatory patterns before they are mechanized on a large scale. Additionally, they seek to embed democratic values such as equality, non-discrimination, and transparency as engineering requirements from the initial design phase rather than as ethical aspirations addressed at the end. Explainable AI (XAI) enables civil society and independent audit bodies to scrutinize on an ongoing basis how demographic groups are classified and treated, thereby rendering algorithmic accountability a continuous democratic check rather than a retrospective exercise (Marque et al. 2024; Maeng 2024).

The third group-level recommendation confronts the most politically contentious threat posed by AI: labor market restructuring that concentrates economic gains among a narrow elite while displacing the workers whose labor made that progress possible. It is evident that economic security serves as the foundational element for political participation. Consequently, individuals who have experienced a diminution in income and bargaining power concurrently undergo a material reduction in their capacity for self-rule. It is imperative that policy intervene at two distinct points in order to effectively address this issue. First, it is incumbent upon governments to incentivize corporations to design AI systems around the concept of human-AI complementarity. This entails the augmentation and reallocation of human tasks, rather than the wholesale displacement of human labor. Second, governments must allocate resources to the development of AI literacy and reskilling programs across all occupational sectors, ranging from blue-collar manufacturing to white-collar knowledge work. Concurrently, they should explore redistributive mechanisms, such as a "Risk Tax" on frontier AI development, which could be redirected toward the establishment of universal social safety nets (Elbaum and Mallaby 2026). The fundamental principle is evident: if the gains in productivity achieved by artificial intelligence are collectively derived, drawing upon public research, infrastructure, and the labor of citizens, then the wealth they generate must be democratically distributed.

## Conclusion

This briefing has examined the escalating threats posed by artificial intelligence (AI) to both procedural and substantive democracy, and how these threats are unfolding in South Korea. At the procedural level, meaningful responses are underway: The implementation of AI-based deepfake detection technology in anticipation of the recent local elections signifies an escalating cognizance of the risks posed by artificial intelligence to the integrity of elections. However, when assessed in terms of substantive



democracy, advances remain constrained. Policies addressing the erosion of epistemic agency, the drift toward technocracy, and the deepening of structural inequality remain largely at the stage of discussion. The existence of this discrepancy can be attributed to two structural factors. The development of AI in South Korea has been predominantly driven by industrial interests, which have given precedence to technological advancement over democratic governance. The rapid advancements in AI have outpaced society's ability to effectively regulate their implications, resulting in a discrepancy between the pace of technological evolution and societal oversight. The Republic of Korea AI Action Plan demonstrates a dual inclination, exhibiting strength in promoting the acceleration of AI development and adoption while exhibiting a relative weakness in addressing its democratic implications.

Nevertheless, there are grounds for cautious optimism. In March 2026, the Presidential Council on National Artificial Intelligence Strategy established a new AI Democracy Subcommittee, which was charged with the protection of democratic values and the promotion of inclusive AI. This is a significant development, but it is imperative to distinguish between the concepts of institutional creation and institutional effectiveness. It is incumbent upon the South Korean government to translate the Subcommittee's mandate into binding and enforceable obligations. Failure to do so will result in the well-documented failure mode of constructing sophisticated governance frameworks while permitting industry-driven AI development to proceed unencumbered.

Among the democracies grappling with this challenge, South Korea is uniquely positioned to establish a pioneering model for democratic AI governance. The following three strengths support this assertion: first, the presence of world-class digital infrastructure; second, the existence of a citizenry that is digitally literate and civically engaged; and third, a demonstrated capacity for rapid institutional innovation. In order to develop a governance framework that is technologically sophisticated and democratically legitimate, South Korea should integrate deliberative platforms such as vTaiwan and design inclusive AI-driven public services. By doing so, South Korea will be able to position itself as a credible global leader in a region where leadership is urgently needed.

The ramifications of this military engagement extend well beyond South Korea's territorial boundaries. Across Asia, the emergence of digital authoritarianism—a governance model characterized by the use of AI-powered surveillance, algorithmic social control, and centralized information management as instruments of political domination—suggests a new trend in governance that is increasingly exportable. However, this model currently lacks a democratic alternative that is comparable in scope. It is at this juncture that South Korea's opportunity transcends into a responsibility. Should South Korea successfully develop an "AI for Democracy" framework that is both human-centric and technologically sophisticated, it will demonstrate something of profound regional significance: that technological advancement and democratic values are not in tension, but mutually reinforcing. In a region where that proposition is increasingly contested, South Korea's capacity to embody that demonstration may prove to be its most consequential democratic export.■

## References

- Elbaum, Sebastian and Sebastian Mallaby. 2026. "The AI Trilemma: How to Regulate a Revolutionary Technology." *Foreign Affairs*. February 13. <https://www.foreignaffairs.com/united-states/ai-trilemma>.
- Hankookilbo*. 2026. "1,600 Illegal Deepfake Campaign Violations Detected in Jeonnam-Gwangju (In Korean)." March 31. <https://www.hankookilbo.com/news/article/amp/A2026033111070004376>.
- Hong, Seokhan. 2024. "The Risks of Artificial Intelligence on Elections and Normative Responses (In Korean)." *Public Law* 53 (2): 185–213.
- International IDEA. 2026. *Harnessing Artificial Intelligence to Enhance Political Finance Oversight*. Stockholm: International IDEA.
- Jackson, Dean and Samuel Woolley. 2025. "AI's Real Dangers For Democracy." *Journal of Democracy* 36 (4): 139–150.
- Jungherr, Andreas. 2023. "Artificial Intelligence and Democracy: A Conceptual Framework." *Social Media + Society* 9 (3): 1–14.
- König, Pascal D. 2023. "Citizen Conceptions of Democracy and Support for Artificial Intelligence in Government and Politics." *European Journal of Political Research* 62 (4): 1280–1300.
- Marques, Marta Sofia, Maria Anastasiadou, and Vitor Santos. 2024. "Framework for the Application of Explainable Artificial Intelligence Techniques in the Service of Democracy." *Transforming Government: People, Process and Policy* 18 (4): 638–656.
- National Election Committee. 2026. "Key Contents of the Partial Amendment to Political Relations Laws (Passed by Plenary Session, December 20, 2023): Provisions on Deepfakes, Women's Candidate Recommendation Subsidies, and Related Matters (In Korean)." <https://www.nec.go.kr/site/nec/ex/bbs/View.do?cbIdx=1130&bcIdx=196646>.
- Seoul Metropolitan Government. 2025. 2025 Seoul Survey. Seoul: Seoul Metropolitan Government.
- The Presidential Council on National Artificial Intelligence Strategy. 2026. Republic of Korea AI Action Plan (2026–2028). Seoul: The Presidential Council on National Artificial Intelligence Strategy.
- Yang, JiSoo. 2026. "AI-Driven Consensus Formation and High-Conflict Policy Governance: An Analysis of vTaiwan's Mechanisms (In Korean)." *The Korean Journal of Political Science* 34 (1): 1–25.



*Yonhap News*. 2026. "Detecting Deepfakes in Local Elections: AI System Identifies Even Generative AI Manipulation (In Korean)." March 10.  
<https://www.yna.co.kr/view/AKR20260310090151530>.

Yun, Soo-Jeong. 2021. "Artificial Intelligence and Democracy (In Korean)." *Journal of Constitutional Court Research* 8 (2): 3–27.

- **Sunghack Lim** is a Professor at the University of Seoul.

The East Asia Institute takes no institutional position on policy issues and has no affiliation with the Korean government. All statements of fact and expressions of opinion contained in its publications are the sole responsibility of the author or authors.

This program was funded in part by the National Endowment for Democracy (NED).

“Harnessing AI to Strengthen South Korean Democracy”

979-11-7584-080-5 95340

Date of Issue: 15 April, 2026

Edited by Jaehyun Im

For inquiries:

Jaehyun Im, Research Associate

Tel. 82 2 2277 0746 (ext. 209)      jhim@eai.or.kr

The East Asia Institute  
1 Sajik-ro 7-gil, Jongno-gu, Seoul 03028, Republic of Korea  
Phone 82 2 2277 1683 Fax 82 2 2277 1684