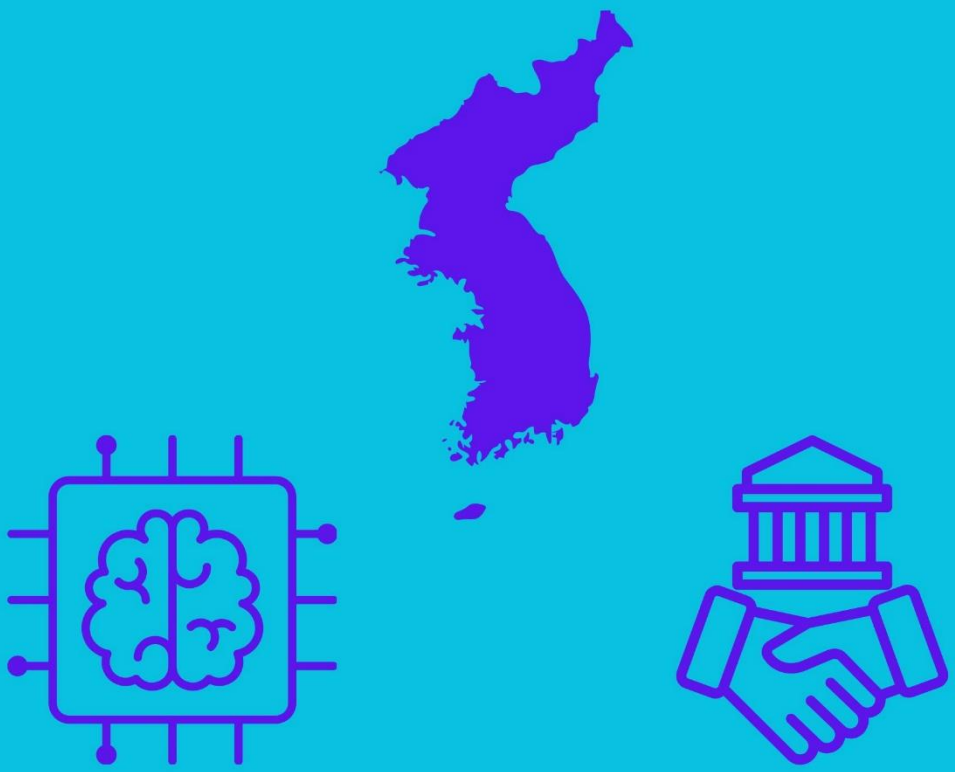
Asia Democracy Research Network

South Korea's Approach to **Democratic AI Governance**

Asia Democracy Research Network

July 2026



In 2026, the Asia Democracy Research Network (ADRN) selected "AI and Democratic Governance" as the subject of a joint research project, with South Korea serving as a dedicated country case study authored by the Korean member research team of ADRN.

Building on ADRN's broader inquiry into democratic values in AI governance, this series of issue briefings sets out to trace South Korea's institutional, societal, and international efforts to govern artificial intelligence, and to assess how effectively these efforts safeguard and strengthen democratic accountability amid both AI-driven risks and opportunities.

The report investigates contemporary questions such as:

- How has South Korea developed its institutional architecture for AI governance under the AI Basic Act, and how do citizen-participatory mechanisms fit within its "AI Basic Society" vision?
- What threats does AI pose to procedural democracy (deepfake-driven electoral manipulation) and substantive democracy (epistemic erosion, structural inequality), and how has South Korea responded?
- How is South Korea's digital government and AI governance model diffused to developing countries through official development assistance?
- To what extent is South Korea's "Sovereign AI" model democratically legitimate, and could it be transferred to other middle-power democracies?

Drawing on a rich array of resources and analyses—including a review of the 2025 Artificial Intelligence Basic Act and municipal AI guidelines, a nationally representative public survey, and a specialized survey of civil servants—the report provides institution-specific and demographic insights, highlights emerging gaps between AI adoption and governance readiness, and suggests policy directions to strengthen democratic accountability in Korea's evolving AI landscape.

“South Korea’s Approach to Democratic AI Governance”
ISBN (electronic) 979-11-7584-119-2 95340

Table of Contents

Abstract

Part 1
South Korea's Approach to AI Governance for Democracy
Sook Jong Lee, East Asia Institute; Sungkyunkwan University

Part 2
Harnessing AI to Strengthen South Korean Democracy
Sunghack Lim, University of Seoul

Part 3
AI Aid for Good Governance: Reconfiguring South Korea's International Development Cooperation in the AI Era
Myung-jae Moon, Yonsei University

Part 4

South Korea's Sovereign AI Initiative for Democracy
: A Transferable Model for the Global Community
Jong Jin Lee, Institute for Peace and Unification Studies

Abstract

South Korea's rapid digital transformation, matched by sustained democratic consolidation since the late 1980s, offers a distinctive case for examining how AI governance frameworks are legislated, how threats to democracy are identified and countered, how such models are diffused abroad, and how a distinctive AI strategy is built for international transfer. This issue briefing series, sponsored by the Asia Democracy Research Network (ADRN), synthesizes four interrelated bodies of conceptual and policy analysis. Part I analyzes the institutional architecture of AI governance under the AI Basic Act (2024/2026) and citizen-participatory mechanisms within the "AI Basic Society" vision. Part II examines threats to procedural democracy (deepfake-driven electoral manipulation) and substantive democracy (epistemic erosion, structural inequality), alongside policy responses. Part III traces the diffusion of Korea's digital government model through KOICA's development assistance under the "AI for All Global Community" initiative. Part IV assesses the architecture and transferability of Korea's "Sovereign AI" model as a middle-power alternative to Silicon Valley techno-liberalism and Chinese state-led AI authoritarianism. Together, the four parts constitute evidence-based reflections on the development of artificial intelligence (AI) and democratic governance in South Korea.

Part I: Institutional Architecture of AI Governance and Citizen-Participatory Public Forums

Part I traces South Korea's shift from voluntary self-regulation under the 2020 National Guidelines for AI Ethics to the legally binding Framework Act on the Development of Artificial Intelligence and the Establishment of Trust ("AI Basic Law"), adopted in December 2024 and promulgated in January 2025 (Lee, 2026). The law establishes a Presidential Committee as a national AI "control tower," an AI Safety Institute, and a triennial AI Master Plan requirement, and imposes watermarking and safety obligations on high-impact AI systems along with extraterritorial application to foreign operators. The Lee Jae-myung administration has advanced "AI Basic Society" (ABS) as its overarching policy vision, promoting AI accessibility, government-led AI public forums, and deliberative mechanisms such as Digital Citizens' Assemblies, though concrete procedural safeguards for legitimacy remain underdeveloped. Internationally, the AI Seoul Summit (2024), the Korea-Singapore AI Cooperation Framework, and linkages with the Japan-India AI Initiative reflect Korea's search for a "middle-power alternative" distinct from the US market-driven, EU rights-based, and Chinese state-led models.

Keywords: AI Basic Law; AI Basic Society; AI public forum; international AI cooperation

Part II: Threats to Procedural and Substantive Democracy and Policy Responses

Part II applies Jungherr's (2023) analytical framework to map AI's impact on democracy across the individual (self-rule), group (equality), and institutional (elections) levels, distinguishing threats to procedural and substantive democracy (Lim, 2026). At the procedural level, deepfake-enabled electoral interference—exemplified by 1,600 illicit deepfake campaign violations detected in the Jeonnam-Gwangju regional primary—and opaque political finance ("donation splitting") are identified as central threats, with recommendations for a "Public-Interest AI Defense Infrastructure," AI-based political finance auditing, and an "Electoral Integrity AI Model." At the substantive level, the briefing highlights the erosion of epistemic autonomy through algorithmic filter bubbles and technocratic drift, alongside AI-driven structural inequality from biased

training data and labor market disruption, proposing Taiwan's vTaiwan (Pol.is) platform, ex-ante human rights audits, and human-AI complementary labor design as remedies. The AI Democracy Subcommittee established in March 2026 marks institutional progress, though its effectiveness depends on future implementation.

Keywords: procedural/ substantive democracy; deepfake electoral interference; technocracy

Part III: Civil Servant AI Adoption and Governance Readiness

Part III analyzes how South Korea's e-government ODA—having progressed through introduction (1991–2000), growth (2001–2010), and maturity (2011–early 2020s) stages—is being reoriented for the AI era (Moon, 2026). KOICA's public governance ODA share rose from 16.0% in 2020 to 20.2% in 2023, with ICT allocations expanding in parallel, and the Fourth Comprehensive Basic Plan for International Development Cooperation (2026–2030), adopted in February 2026, signals a shift from quantitative expansion to qualitative enhancement. Linked to the Lee administration's AI transformation strategy (National AI Computing Center, data ecosystem expansion, AI talent cultivation), KOICA has formulated a sovereign-AI-based mid- to long-term strategy and expanded "AI-driven innovative ODA," including AI-based water resource management in Indonesia and AI workforce development in Bangladesh. The briefing stresses that translating such support into genuine good governance requires careful assessment of partner countries' institutional readiness—political commitment, data governance, cybersecurity, and accountability mechanisms.

Keywords: digital government ODA; KOICA; AI for All Global Community

Part IV: Sovereign AI and the International Transferability of Democratic Alignment

Part IV traces Korea's construction of "Sovereign AI as a democratic institution"—positioned as a third path between Silicon Valley techno-liberalism and Chinese state-AI authoritarianism—across three phases: ethical foundation (2016–2020), institutional architecture (2020–2025), and the K-AI Initiative (2025–present) (Lee, 2026). It details the Korea AI Safety Institute's three-dimensional risk-mapping methodology (MECI), the K-AI Initiative's narrowing from five to three corporate consortia (with final evaluation due December 2026), and clearly distinguishes codified legal provisions from proposed concepts such as "Democratic RLHF" (citizen-jury-based alignment), "procedural delay," and a three-layer civic oversight structure (red teams, self-regulatory ethics committees, auditable sandboxes). Internationally, the May 2026 launch of the Global AI Hub—joining nine UN agencies and five multilateral development banks—illustrates Korea's role as a "norm exporter" and convener of horizontal partnerships, and the briefing outlines a modular (immediate/medium/long-term) adoption pathway premised on institutional preconditions such as civil society capacity, rule of law, and technical infrastructure.

Keywords: Sovereign AI; democratic alignment; Democratic RLHF; middle-power norm diplomacy

Part 1

South Korea's Approach to AI Governance for Democracy

Sook Jong Lee
Distinguished Fellow, East Asia Institute;
Distinguished Professor, Sungkyunkwan University

Ⅰ. Introduction

The disruptive impacts of AI on democracy range from deepfakes and cognitive outsourcing to electoral manipulation in democratic systems, as well as panopticon-like surveillance in authoritarian regimes. Ovadya argues that dystopias of autocratic centralization and ungovernable decentralization can be avoided through "democratic centralization and democratic decentralization," enabled by four innovations: mini-public deliberation, AI-augmented decision-making, third-party participatory governance, and platform democracy.[1]

While this approach—seeking to advance both AI innovation and democratic governance—is reasonable, much of the current debate on AI's impact on democracy focuses primarily on procedural dimensions, often overlooking AI's potential to link citizen demands to public services. Significant room remains for AI-assisted government innovation to improve service delivery. If implemented wisely, AI-

[1] Aviv Ovadya. 2023. "Reimagining Democracy for AI." *Journal of Democracy* 34 (4): pp. 162-170.

assisted public governance could help narrow service gaps in an era of widening wealth inequality. This briefing examines how South Korea is responding to public demands for both substantive democracy and legitimate procedural democracy.

Ⅱ. South Korea's Recent AI Policy Initiatives

The AI Basic Law establishes a comprehensive regulatory framework that both promotes the development of the AI industry and strengthens safety, ethical standards, and user protections, particularly for high-impact AI systems.

South Korea passed the Framework Act on the Development of Artificial Intelligence and the Establishment of Trust—commonly referred to as the AI Basic Law. The law was adopted by the National Assembly on December 26, 2024, formally promulgated on January 21, 2025, and entered into force on January 22, 2026.

The AI Basic Law establishes a comprehensive regulatory framework that both promotes the development of the AI industry and strengthens safety, ethical standards, and user protections, particularly for high-impact AI systems. It provides for a Presidential Committee to serve as a national AI "control tower," a dedicated AI Safety Institute, and requires the government to formulate an AI Master Plan every three years to enhance national competitiveness. Under the Act, AI-generated content must be clearly labeled—including through watermarks—and developers and operators of "high-impact" AI systems must implement appropriate safety measures, alongside industry development measures such as R&D funding, training data construction, and the designation of AI clusters.

The law's regulatory scope extends to foreign AI systems affecting users or markets within South Korea. Foreign AI companies without a domestic presence must designate a local agent if they meet certain thresholds—such as having more than one million average daily users in Korea or generating substantial domestic revenue. The Ministry of Science and ICT (MSIT) is empowered to investigate violations and may issue corrective orders or impose administrative fines of up to KRW 30 million (approximately USD 22,000).

While comprehensive and ambitious, the Action Plan gives comparatively less attention to democracy than to AI industry development and public-sector AI transformation (AX), yet it still reflects a distinctive social vision for AI.

This marks a shift from voluntary guide lines to mandatory compliance with AI ethics. The previous "National Guidelines for Artificial Intelligence (AI) Ethics," adopted in December 2020 by the Presidential Committee on the Fourth Industrial Revolution, rested on three basic principles: respect for human dignity, the common good of society, and the proper use of technology.[2] These guidelines relied on a safeguard and self-regulation mechanism, deliberately avoiding rigid legal norms deemed inadequate for a fast-changing technological environment. Concerns about excessive regulation hampering industrial growth were also a factor. With the AI Basic Law, however, South Korea became the first country to enforce binding regulations on high-impact AI, while the European Union's AI Act continues to delay its full implementation.

Since President Lee Jae-myung took office in May 2025, South Korea's AI policy has pursued two goals: fostering an internationally competitive AI industry and promoting trust in an AI-based society. The Presidential Committee on AI National Strategy, launched on September 8, 2025 with eight divisions, holds formal decision-making and implementation authority—a significant upgrade from the advisory-only committee under the previous administration. In December 2025, it announced a preliminary ROK AI Action Plan comprising 12 strategic areas and 98 specific tasks, with a final version expected in 2026. While comprehensive and ambitious, the Action Plan gives comparatively less attention to democracy than to AI industry development and public-sector AI transformation (AX), yet it still reflects a distinctive social vision for AI.

[2] AI Ethics Communication Channel. N.d. "The National Guidelines for AI Ethics." https://ai.kisdi.re.kr/eng/main/contents.do?menuNo=500011.

Ⅲ. AI Basic Society

South Korea's approach to AI governance employs terms such as trust, transparency, accountability, and inclusiveness rather than "democracy" per se, reflecting a particular framing of AI governance values. How countries govern AI to protect or innovate democracy naturally varies.

"AI Basic Society" (ABS) is President Lee's overarching AI policy vision: an inclusive digital capacity society in which all citizens—regardless of age, occupation, region, or socioeconomic background—are able to understand and utilize AI. The Action Plan describes ABS as a new social compact, committing the government to build an "AI for All" platform that enables citizens to easily understand AI and AI-driven change, directly experience its applications, and naturally integrate AI into their daily lives. A Nationwide AI Capability Transformation Project is also proposed to provide standard models of basic AI competencies and ensure equitable access to AI learning opportunities.[3]

The Action Plan describes ABS as a new social compact, committing the government to build an "AI for All" platform that enables citizens to easily understand AI and AI-driven change, directly experience its applications, and naturally integrate AI into their daily lives.

Literally, a "Basic Society" is one where citizens' basic livelihoods are guaranteed through rights to social services and strengthened social safety nets.[4] In this context, ABS is a concept emphasizing equitable access to AI-assisted social services. The Action Plan allocates 16 tasks across 24 government offices under this segment. A 'Comprehensive Plan for an AI-Based Basic Society for All,' to be finalized by the second quarter of 2026, "shall aim to enhance citizens' quality of life and strengthen the social safety net, while integrally incorporating the principles of AI democracy, technological ethics, and social inclusion" across seven key sectors: labor, welfare, education, finance, culture, safety, and environment.[5] Implementation measures include: institutionalizing the legal status of public-interest data and AI; allocating 10–

[3] Presidential Committee on National AI Strategy. 2025. "ROK AI Action Plan." December. p. 127.

[4] Lee, Jae-myung. 2025. "People Sovereignty Government, Basic Income or Basic Society." *Le Monde Diplomatique* Korean Edition. August 5. https://www.ilemonde.com/news/articleView.html?idxno=21018. (Accessed: February 10, 2026)

[5] Presidential Committee on National AI Strategy. 2025. "ROK AI Action Plan." December. p. 129.

20% of public-sector AX budgets to public-interest AI; setting minimum private-sector participation levels for social innovation; linking government programs to AI social ventures; building public–private–academic data governance; and establishing ethical review procedures for public-interest data.[6]

Unlike traditional participatory mechanisms with high legal and logistical thresholds, AI-assisted forums can be created with minimal oversight, making it harder for citizens to trust the growing number of such spaces.

The MSIT will launch pilot projects for the AI-based basic society and identify necessary legal and institutional reforms by the fourth quarter of 2027, with nationwide services in AI democracy, inclusive finance, and climate risk management to be initiated by the first quarter of 2030. This process will also promote the securing and open access of public-interest data and the establishment of participatory governance mechanisms.

Ⅳ. AI-Assisted Public Sphere

The South Korean government has also proposed creating an AI-assisted public sphere to strengthen representation and help legitimize procedural democracy in the AI era. Kreps and Kriner identify AI threats to democracy in three key areas: representation, accountability, and trust.[7] Altman highlights risk to direct democracy, arguing that AI-driven astroturfing can distort grassroots movements and undermine deliberation in public forums. Unlike traditional participatory mechanisms with high legal and logistical thresholds, AI-assisted forums can be created with minimal oversight, making it harder for citizens to trust the growing number of such spaces.[8] Government-led AI public forums designed with robust trustworthiness safeguards are therefore worthy of serious exploration.

The Action Plan proposes creating AI public forums and experimental citizen labs. It writes: "The pan-governmental AI public forum shall function as a multi-stakeholder collaborative platform that collects data on the risks and

[6] Presidential Committee on National AI Strategy. 2025. "ROK AI Action Plan." December. p. 130.

[7] Kreps, Sarah and Doug Kriner. 2023. "How AI Threatens Democracy." Journal of Democracy 34 (4): pp. 122-131.

[8] Altman, David. 2026. "The AI Democracy Dilemma." *Journal of Democracy* 37 (1): pp. 32-44.

At the same time, they must be carefully designed to address potential representation gaps among participating citizens, particularly along partisan and generational lines.

opportunities arising from AI transformation and identifies key agenda items for discussion. At the same time, it shall enable experts, stakeholders, and government officials to deliberate together with citizens and develop these discussions into concrete policy measures."[9]

South Korea has a strong track record of digital participatory mechanisms—including opinion surveys, new policy initiatives, evaluations, citizen ombudsman systems, and petitions. Government-led AI forums could fundamentally transform this model of citizen engagement: they can be faster and more effective through access to public data and multi-stakeholder participation within a shared digital space. At the same time, they must be carefully designed to address potential representation gaps among participating citizens, particularly along partisan and generational lines. Designating non-partisan NGOs as the coordinating "control tower" of these public forums—supported by government infrastructure and public data—could be a viable approach.

Ⅳ. International Cooperation for AI Governance

No global AI governance framework yet exists despite extensive discussions on AI's benefits and harms. In September 2024, the UN High-Level Advisory Body on AI released its final report, Governing AI for Humanity, recommending seven items as an agile and flexible governing approach.[10] So far, little progress has been made. AI governance structures remain fragmented across different forums, lacking the coherence required to manage rapid AI development and deployment.

---

[9] Presidential Committee on National AI Strategy. 2025. "ROK AI Action Plan." December. p. 132.

[10] Seven recommendations are: the establishment of an IPCC-like independent international scientific panel on AI to provide authoritative, science-based assessments; the launch of a twice-yearly intergovernmental and multi-stakeholder policy dialogue on AI governance to share best practices and promote common understanding on the implementation of AI governance; the creation of an AI standard exchange bringing together representatives form national and international standard-development organizations, tech companies, and representatives from civil society and the international science panel; the creation of an AI development network to align regional and global AI capacity efforts and build AI governance capacity of public officials; the creation of a global AI fund to put a floor under the AI divide; and the creation of a global AI date framework outlining data-related definitions and principles for global governance of AI training date; and the creation of AI office within the Secretariat. United Nations. N.d. "Governing AI for Humanity: Final Report." https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

The central challenge is therefore not to choose among the US, EU, or Chinese models, but to develop an alternative framework that reconciles technological access with robust safeguards.

Three dominant models currently shape the global landscape. The United States advances a market-driven model led by private innovation and corporate self-regulation, reflecting its broader market-oriented regulatory philosophy. The European Union has adopted a comprehensive regulatory approach grounded in human rights and democratic values, exemplified by the world's first comprehensive AI Act introduced in 2024. China, by contrast, follows a government-led model rooted in techno-nationalism, prioritizing political stability and social control while regulating AI in targeted sectors and use cases.

For many Asian democracies, none of these models offers a perfect fit. Unlike the US or China, they do not host dominant global AI firms or command the capital to independently build full AI ecosystems. At the same time, they are eager to develop domestic AI capabilities and secure access to advanced technologies, yet—similar to the EU—remain highly sensitive to the societal risks associated with AI deployment. This combination of technological aspiration and risk awareness places them in a structurally distinct position. The central challenge is therefore not to choose among the US, EU, or Chinese models, but to develop an alternative framework that reconciles technological access with robust safeguards. In the Indo-Pacific, this requires transforming existing guidelines into institutionalized cooperation, strengthening regional dialogue, and aligning with global governance processes.

1. AI Diplomacy

As a leading semiconductor manufacturer and advanced digital society, South Korea has actively joined global AI cooperation. It hosted the AI Seoul Summit in May 2024, focused on safety, innovation, and inclusivity. The Summit's Declaration recognized the importance of interoperability between AI governance frameworks and called for creating or expanding AI safety institutes, research programs, and supervisory bodies. It also called for protecting democratic values, the rule of law, human rights, and fundamental freedoms, and for bridging AI and digital divides

While AI cooperation has become a common element in international partnerships, South Korea needs to identify more focused areas for collaboration, as priorities will inevitably vary depending on partner countries' levels of economic development and digital infrastructure.

between and within countries to advance the UN Sustainable Development Goals.[11]

South Korea has also begun incorporating AI collaboration into its bilateral partnerships. During President Lee's state visit to Singapore on March 2, 2026, both governments shared the ABS vision and expressed intent to pursue an "AI Cooperation Framework." While AI cooperation has become a common element in international partnerships, South Korea needs to identify more focused areas for collaboration, as priorities will inevitably vary depending on partner countries' levels of economic development and digital infrastructure.

The India–Japan AI Initiative offers a useful model in this regard. Strengthened through the Japan–India AI Strategic Dialogue launched in January 2026, it focuses on creating a trustworthy AI ecosystem, advancing large language models (LLMs), and supporting startups to counter technology monopolies and promote safe and ethical AI use. At its core is the joint development of international standards for "Trustworthy AI," based on principles such as personal data protection, algorithmic transparency, and the elimination of bias—distinct from the approaches taken by US Big Tech or China. The initiative also carries geopolitical significance as a potential "third AI bloc" amid intensifying technological competition.[12]

As Japan advances AI ethics and democratic norms—as reflected in the Hiroshima Process, which established international guiding principles for AI-developing organizations—South Korea should enrich the ABS concept with a

---

[11] Ministry of Foreign Affairs, Republic of Korea. 2024. "Seoul Declaration for Safe, Innovative and Inclusive AI by Participants Attending the Leaders' Session of the AI Seoul Summit." May 21. https://www.mofa.go.kr/eng/brd/m_5674/view.do?seq=321007. The Summit also stated to strengthen international cooperation on Al governance through engagement with other international initiatives at the UN and its bodies, G7, G20, the Organization for Economic Co-operation and Development (OECD), the Council of Europe, and the Global Partnership on AI (GPAI). The Hiroshima AI Process Friends Group, the updated OECD AI principles, and the recently adopted UN General Assembly resolution "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development," the Global Digital Compact in advance of the Summit of the Future in September 2024 were all acknowledged.

[12] Kurihara, Toshihiko and Sukirt Kaur. 2026. "Indo–Japanese Collaboration on Artificial Intelligence." Center for Social and Economic Progress, *Executive Policy Brief*. January. https://csep.org/wp-content/uploads/2026/01/INDO%E2%80%93JAPANESE-COLLABORATION-ON-ARTIFICIAL-INTELLIGENCE.pdf.

concrete operational framework that goes beyond broad vision. AI-related foreign assistance programs could serve as an important avenue for operationalizing this framework. South Korea also needs to step up AI cooperation with ASEAN. In 2025, ASEAN expanded its AI Governance and Ethics Guide to include nine recommendations addressing the risks of generative AI, positioning itself as a potential regional norm entrepreneur.[13] However, for this framework to become a viable governance model, it must be embedded in broader regional cooperation involving major AI players such as Korea, Japan, and Taiwan. With its strengths in AI infrastructure and hardware, South Korea is well positioned to help fill this gap.

2. AI Foreign Assistance

The Action Plan envisions a Global Alliance for an AI Basic Society—an open collaborative platform bringing together government, industry, academia, and civil society—to link international standards on AI ethics, trustworthiness, and fairness; build inclusive data governance; expand public AI capabilities; and promote global joint projects. The Alliance "aims to contribute to Korea's emergence as a global hub for the AI Basic Society model…a strategic diplomatic initiative to present a Korean model of 'K-democracy and growth in the AI era' to the international community by exploring an AI development strategy that is democratic, public-interest-oriented, and growth-driven."[14]

However, for this framework to become a viable governance model, it must be embedded in broader regional cooperation involving major AI players such as Korea, Japan, and Taiwan.

To translate this vision into concrete programs, KOICA concluded an MOU with the National Information Society Agency (NIA) in February 2026 to provide "AI for All" ODA. Building on prior cooperation—including the dispatch of IT volunteer corps, consulting on development cooperation projects, and expert advisory services—the revised agreement significantly expands the scope to include the AI field. The two institutions plan to cooperate in: promoting AI ethics in international development cooperation; providing expert advisory support on AI

[13] ASEAN Secretariat. 2025. "Expanded Guide: AI Governance and Ethics – Generative AI." https://asean.org/wp-content/uploads/2025/01/Expanded-ASEAN-Guide-on-AI-Governance-and-Ethics-Generative-AI.pdf.

[14] Presidential Committee on National AI Strategy. 2025. "ROK AI Action Plan." December. p. 135.

policies and regulatory frameworks for developing countries; sharing domestic and international networks and information; and promoting institutional publications.

To make these plans effective, South Korea's global vision must translate into concrete democratic models that can adapt to diverse local contexts. Because democratic institutions in many developing countries do not function effectively, clear guardrails are needed to ensure that AI-focused ODA genuinely supports democratic governance. At the same time, the idea of "K-democracy in the AI era" remains underdefined. Treating AI merely as an extension of South Korea's digitally active civic culture risks overlooking its more transformative—and potentially disruptive—impact in the Global South.

Ⅴ. Conclusion

To make these plans effective, South Korea's global vision must translate into concrete democratic models that can adapt to diverse local contexts.

Shany's proposed AI Bill of Rights outlines seven key rights: the right of access to AI systems; the right to privacy-related protections from harmful AI uses; the right to be free from algorithmic bias and unfairness; the right to algorithmic transparency and explainability; the right not to be subject to algorithmic manipulation; the right to human decision-making and human-to-human interaction; and the right to accountability for harms caused by the use of AI systems.[15]

South Korea's "AI for All" vision aligns closely with the right to access AI systems, though issues of algorithmic transparency and accountability remain less explored. The country's strengths lie in reducing AI access gaps and enhancing public service effectiveness. Building on its relatively successful digital governance, which has already enabled citizen participation in policy processes, South Korea is now rapidly expanding AI-assisted services in citizens' interactions with government.

In a context of increasing political polarization, AI-assisted public forums could also produce negative effects by intensifying political confrontation. At the

[15] Shany, Yuval. 2025. "White Paper: The Need for and Feasibility of an International AI Bill of Human Rights." Institute for Ethics in AI, University of Oxford. November. https://afp.oxford-aiethics.ox.ac.uk/sitefiles/white-paper-professor-yuval-shany.pdf.

same time, AI platforms designed to broaden participation could help depolarize public discourse and contribute to solving social problems. Efforts to strengthen the deliberative capacity of citizens should therefore involve not only the government but also citizens themselves and other private-sector actors.

South Korea's international cooperation for responsible AI also deserves attention. While its multilateral initiatives are expanding, they could be complemented by more regionally focused cooperation. Collaboration with Japan, Singapore, and Taiwan could be particularly promising. When engaging with the Global South, South Korea's AI-related ODA should also clarify how AI assistance can support democratic institutions—moving beyond technical cooperation and training exchanges.

Part 2

Harnessing AI to Strengthen South Korean Democracy - Evidences from a National Survey

Sunghack Lim
Professor, University of Seoul

Ⅰ. 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 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.

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.

Ⅱ. 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).

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).

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.

Ⅲ. 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).

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 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.

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.

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.

Ⅳ. 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).

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.

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 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 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

The technological capacity that renders AI a potential threat to substantive democracy can, under different institutional conditions, become a resource for strengthening democracy.

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.

To address these concerns, governments must enforce algorithmic transparency requirements and mandate regular, independent external audits of major digital platforms.

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).

With regard to the matter of timing, governments must transition from a post-hoc detection model to an ex-ante human rights audit model.

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.

Ⅴ. Conclusion

The development of AI in South Korea has been predominantly driven by industrial interests, which have given precedence to technological advancement over democratic governance.

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.

Among the democracies grappling with this challenge, South Korea is uniquely positioned to establish a pioneering model for democratic AI governance.

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.

Part 3

AI Aid for Good Governance – Reconfiguring South Korea's International Development Cooperation in the AI Era

Myung-jae Moon
Distinguished Professor, Yonsei University

Ⅰ. Introduction

The rapid advancement of digital technologies has introduced new administrative and technical demands while prompting a fundamental reassessment of governance capacity in the public sector. As digital transformation accelerates—particularly with the proliferation of artificial intelligence (AI)—governments face both significant opportunities and emerging risks. Digital government is increasingly defined by the integration and application of artificial intelligence across public administration. As delineated in Table 1, a diverse array of AI technologies, encompassing automation and algorithmic systems, is being implemented across fundamental governmental functions, ranging from internal administration and policymaking to service delivery and oversight. These applications have the potential to enhance key public values associated with good governance, including efficiency, effectiveness, responsiveness, accountability, transparency, and participation (OECD 2025).

Table 1. Understanding the Use of AI in Governments

<table>
<tr><th>AI Tasks</th><th>Government Activities</th><th>Opportunity Area (Administrative Values)</th></tr>
<tr><td rowspan="4">Recognition, Event detection, Forecasting, Personalization, Interaction support, Goal-driven Optimization, Content generation, Reasoning with knowledge</td><td>Internal Operations</td><td rowspan="2">Efficiency
Effectiveness</td></tr>
<tr><td>Policymaking (Decision-making)</td></tr>
<tr><td>Public Service Delivery to Citizens and Businesses</td><td rowspan="2">Responsiveness
Participation</td></tr>
<tr><td>Internal Oversight, External Oversights</td><td rowspan="2">Accountability
Transparency</td></tr>
</table>

Modified based on Governing with Artificial Intelligence by OECD (2025) p. 23.

Despite this promise, governments face substantial challenges in managing AI-related risks. Public institutions frequently encounter challenges in maintaining pace with rapid technological advancements and in mitigating the potential adverse consequences thereof. These challenges are particularly acute in developing countries, where vulnerability to technological disruption is higher and administrative and institutional capacities are more limited. This has the potential to constrain governance performance and democratic resilience.

It has been posited that South Korea's economic development model was once characterized as state-led. However, the country has since been recognized by many as having achieved economic growth and democratic consolidation. This transformation has been supported by strong administrative capacity, participatory governance, and high-quality public service delivery, reinforced by sustained reform and effective digital adoption. The digital government of South Korea has garnered international acclaim, as evidenced by its inclusion in the United Nations' Global E-Government Survey and OECD evaluations. In recent developments, the concept has evolved towards an AI-augmented governance model, emphasizing accountability and transparency, thus positioning itself as a paradigm for public sector innovation.

In this regard, South Korea has emerged as a leading example of good governance, demonstrating how information and communication technologies (ICTs) can be effectively leveraged to enhance the quality, transparency, and participatory nature of public administration.

This issue brief explores the manner in which the South Korean government fosters good governance and fortifies institutional capacity in developing countries. It accomplishes this by aligning its national AI agenda with digital government initiatives under official development assistance (ODA) programs. The study commences with an examination of contemporary trends in South Korea's digital government ODA. The analysis then proceeds to examine recent national AI policies and the strategic alignment of ODA with global agendas such as "AI for good governance" and "AI for all." The focus of this analysis is on enhancing efficiency, transparency, and accountability while building AI capacity in developing countries. The study ultimately provides policy recommendations for the advancement of Korean-style ODA (K-ODA) in the AI era.

Ⅱ. State of South Korea's ODA Programs in Digital Technologies and the Emergence of AI Government

The advent of the digital government in South Korea can be traced back to 1967, marking the inaugural introduction of computers for the purpose of processing population census data (MOIS 2017). Since that time, South Korea has achieved substantial progress in administrative efficiency and the development of advanced public service systems for citizens and businesses. This progress has occurred alongside the country's rapid and compressed economic development. It is noteworthy that these initiatives were sustained as a strategic priority even during periods of economic adversity, such as the financial crisis of the late 1990s. A notable aspect of South Korea's digital government development is its integration within a cohesive presidential agenda, situated within the broader context of government-wide reform initiatives. In this regard, South Korea has emerged as a leading example of good governance, demonstrating how information and communication technologies (ICTs) can be effectively leveraged to enhance the quality, transparency, and participatory nature of public administration.

Consequently, South Korea's digital government has garnered widespread recognition as a paradigm for government reform, eliciting considerable interest from developing countries. Consequently, digital government initiatives have become one of the most popular and prioritized areas within official development assistance (ODA) programs, particularly in the field of public governance. A categorization of South Korea's digital government ODA projects reveals three distinct developmental stages: introduction, growth, and maturity (Kim 2021). The initial stage (1991–2000) placed a primary emphasis on government-to-government (G2G) initiatives, encompassing the establishment of administrative computer networks and the provision of digital equipment and infrastructure in developing countries. The period from 2001 to 2010, known as the growth stage, witnessed a significant expansion of e-government systems, with notable support given to flagship projects such as the electronic procurement system (KoNEPS), the intellectual property administration system (KIPOnet), and the customs clearance system (UNIPASS). Concurrently, broader G2G and government-to-business (G2B) initiatives were also pursued.

Consequently, digital government initiatives have become one of the most popular and prioritized areas within official development assistance (ODA) programs, particularly in the field of public governance.

The maturity stage, spanning from 2011 to the early 2020s, is characterized by two predominant trends: diversification and geographical expansion of project areas. In addition to fortifying digital government infrastructure and public service delivery systems, this phase has seen the incorporation of platforms designed to promote citizen participation. For instance, the National Petition System (e-People), which facilitates citizen engagement and supports anti-corruption efforts, has been widely introduced to developing countries as a model of participatory and accountable digital governance.

Table 2. KOICA's ODA Budget by Sector (2020–2023)

Unit: Million KRW

Sector	2020 (KRW)	Share	2021 (KRW)	Share	2022 (KRW)	Share	2023 (KRW)	Share
Public Governance	108.294	16.00%	128,186	17.60%	142,165	17.70%	193,413	20.20%
Education	134,675	19.80%	100,400	13.80%	110,529	13.70%	148,139	15.50%
Technology, Environment & Energy	71,876	10.60%	107,330	14.70%	127,419	15.80%	137,385	14.40%
Emergency Relief	17,314	2.60%	13,084	1.80%	14,275	1.80%	17,850	1.90%
Agriculture, Forestry & Fisheries	65,094	9.60%	91,071	12.50%	96,998	12.10%	110,607	11.60%
Health & Medical Services	130,176	19.20%	135,360	18.60%	139,215	17.30%	155,253	16.20%
Other	151,216	22.30%	152,336	20.90%	173,867	21.60%	193,244	20.20%
Total	678,647	100%	727,768	100%	804,468	100%	955,891	100%

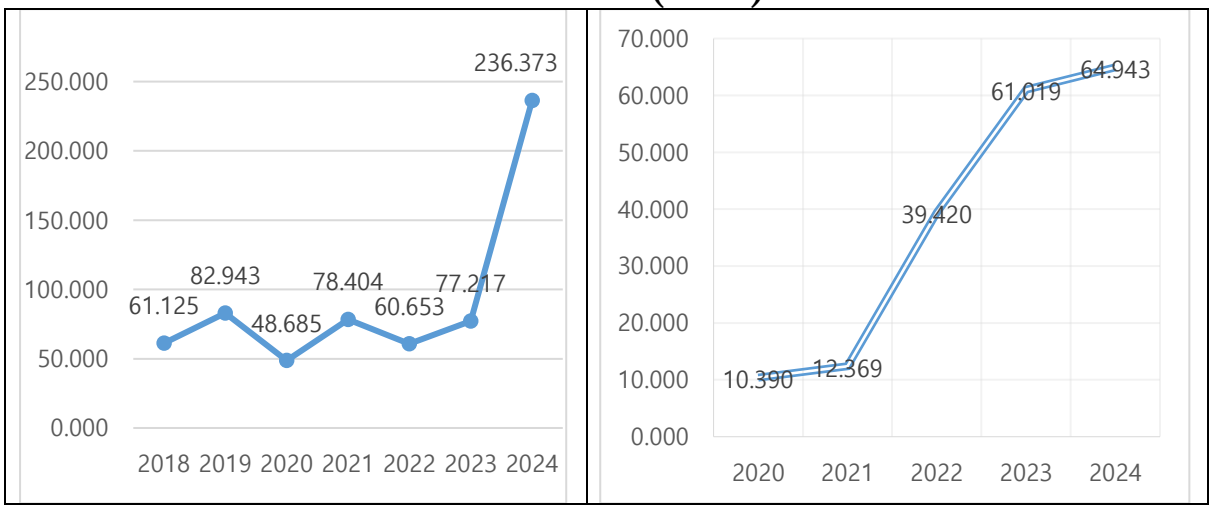
Source: KOICA (2026). https://www.koica.go.kr/koica_kr/901/subview.do accessed on April 15, 2026.

As indicated by the findings presented in Table 2, KOICA's ODA portfolio underwent a substantial expansion between 2020 and 2023, with an increase from 678.647 billion KRW to 955.891 billion KRW. Of particular significance is the shift in the composition of spending, wherein the ODA in Public governance exhibited an increase from 16.0% to 20.2%, while technology, environment, and energy experienced an increase from 10.6% to 14.4%.

This pattern suggests a gradual shift in Korea's ODA, with a transition from traditional social-sector assistance toward governance capacity, digital infrastructure, climate technology, and institutional modernization. This is of particular pertinence to the field of AI for ODA, as the development of AI-related technologies cannot be regarded as a mere technical undertaking. The implementation of AI necessitates

several key elements, including administrative capacity, data systems, digital public infrastructure, ethical governance, human resources, and sectoral applications in health, education, climate, and public administration.

Figure 1. South Korea's ODA Budget for Code 15110 (Public Sector Policy and Administrative Management) and Code 22040 (ICTs)



Made by the author based on the data collected from the OECD Data Explorer (https://data-explorer.oecd.org)

As demonstrated in Figure 1, the OECD Data Explorer indicates that the Republic of Korea (ROK) has notably augmented its budget for Official Development Assistance (ODA) in recent years, allocating significant resources to public sector capacity-building and reform programs, as well as to information and communication technologies (ICTs), classified under Codes 15110 and 22040, respectively. It is important to acknowledge the South Korean government's growing commitment to leveraging information and communication technology (ICT) for the enhancement of governance. This commitment is underscored by the Seoul Declaration of the Second Summit for Democracy in 2023, which places significant emphasis on the imperative of anti-corruption, transparency, institutional integrity, and the rule of law as crucial elements for effective governance and the promotion of democracy (MOFA 2024).

The South Korean government has implemented a series of ICT-integrated projects with the objective of promoting comprehensive and scalable digital development in partner countries. These projects have been implemented across six

key sectors: smart cities, digital government, telemedicine, smart farming, digital education infrastructure, and smart power systems. As the mid-2020s approach, South Korea has endeavored to align its digital government ODA programs more closely with AI-related initiatives. This initiative follows a comprehensive review of its development cooperation portfolio.

The Republic of Korea (ROK) has set its sights on expanding its national "AI for All" initiative to the global stage. In pursuit of this objective, the ROK has identified a strategic priority to enhance AI collaboration with countries in the Global South. The agenda is predicated on the vision of a Global AI Basic Society, with an emphasis on the establishment of an inclusive global AI ecosystem in which the benefits of AI technologies are disseminated extensively and equitably across the global population. It also underscores the significance of global partnerships and the strategic advancement of AI-based international development cooperation.

In pursuit of this objective, the ROK has identified a strategic priority to enhance AI collaboration with countries in the Global South.

These initiatives signify more than a mere technological upgrade; they are indicative of a paradigm shift in the field. These initiatives signal a substantial shift in South Korea's development cooperation priorities, characterized by an emphasis on enhancing the quality of ODA projects, fortifying global collaboration, and promoting inclusive and sustainable growth through the utilization of AI for the benefit of the global community.

Ⅲ. South Korea’s AI Policy and Major Initiatives under the Lee Jae-myung Administration

The administration of President Lee Jae-myung has identified artificial intelligence (AI) as a pivotal element of its national agenda, with the ambitious objective of positioning South Korea among the top three global AI powers (G3). To advance this vision, the government has launched a comprehensive AI-driven transformation strategy to strengthen national competitiveness. This strategy includes 30 flagship initiatives targeting AI adoption and a hyper-innovative economy. The government has also established the Presidential National AI Strategy Committee to coordinate and oversee digital and AI policy across ministries.

The administration's digital policy framework is anchored in two overarching objectives: achieving global AI leadership and fostering an inclusive "AI for All" society. The organization's strategic initiatives are categorized into four distinct areas: infrastructure, data, talent, and governance.

The ongoing reform initiative entails a transition of public data systems to real-time, API-based architectures, along with a shift toward a more proactive, demand-driven data release model.

Firstly, the government is undertaking initiatives to expand its AI infrastructure through public–private collaboration. Key efforts include the planned National AI Computing Center, which is backed by up to KRW 2 trillion in investment and aims to secure computing capacity exceeding one exaflop. The supplementary budget, amounting to KRW 1.46 trillion, is earmarked for the acquisition of GPUs, while an AI Transformation and New Industry Policy Fund with a budget of KRW 810 billion is to be established. These measures are to be accompanied by procurement partnerships with both global and domestic firms.

Secondly, the administration is fortifying the data ecosystem by recognizing data as a pivotal economic asset. The objective of this initiative is to expand the national data market to KRW 50 trillion by 2030. This expansion is to be achieved through three primary strategies: 1) the augmentation of public data access, 2) the integration of support for anonymization, and 3) the development of data trading platforms. The ongoing reform initiative entails a transition of public data systems to real-time, API-based architectures, along with a shift toward a more proactive, demand-driven data release model. This transition is intended to enhance data quality, standardization, and machine readability.

Thirdly, the government is allocating resources to the cultivation and attraction of AI talent. This encompasses the integration of artificial intelligence (AI) and data science education at all levels of education, alongside the implementation of policies designed to attract global talent through financial incentives, institutional flexibility, and measures such as alternative military service options.

The administration has recently initiated a series of legal and institutional reforms with the aim of fostering innovation in the field of artificial intelligence. These reforms are being implemented with a focus on ensuring transparency, accountability, and safety. These efforts concentrate on elucidating eligibility criteria for industry support, establishing regulatory standards for high-impact AI systems,

and constructing comprehensive governance frameworks to manage emerging risks.

Ⅳ. South Korea's New ODA Initiative and "AI for the Global Community"

While the plan provides a strategic foundation for the next five years, debates persist over balancing national interests with global development goals.

As artificial intelligence (AI) emerges as a pivotal catalyst for economic, administrative, and societal transformation, the field of development cooperation must evolve beyond conventional infrastructure and basic information and communication technology (ICT) support. Developing countries are confronted with mounting pressure to fortify their institutional, human, and technological capacities in the domain of artificial intelligence (AI) while concurrently addressing the concomitant ethical, regulatory, and governance challenges.

In February 2026, the Republic of Korea adopted the Fourth Comprehensive Basic Plan for International Development Cooperation (2026–2030), its highest-level ODA framework (Prime Minister's Secretariat 2026). The strategic plan is indicative of two major shifts in the global landscape. Firstly, it reflects an evaluation of the previous plan (2021–2025), under which South Korea nearly doubled its ODA budget—from 3.4 trillion KRW in 2020 to 6.5 trillion KRW in 2025—while strengthening its role as a mid-sized donor.

The new plan signifies a shift from quantitative expansion to qualitative enhancement of ODA. The document delineates four overarching objectives: the promotion of inclusive values, the expansion of mutual benefits, the advancement of innovation, and the establishment of an integrated implementation system. The emphasis is placed on the enhancement of core competencies, the refinement of execution mechanisms, and the augmentation of policy coherence to ensure an increase in overall effectiveness. While the plan provides a strategic foundation for the next five years, debates persist over balancing national interests with global development goals. A fundamental aspect of this initiative entails the promotion of a "Korean-style" development model, which integrates traditional strengths—including health, education, climate, and public administration—with emerging sectors in which South Korea possesses comparative advantages, such as artificial

intelligence and the cultural industry (Prime Minister's Secretariat 2026).

The strategy's core element is the implementation of the inclusive "AI for All" vision within the context of Korean ODA (K-ODA), with the objective of ensuring shared benefits for both South Korea and its partner countries.

In alignment with this overarching framework and the government's comprehensive AI transformation agenda, KOICA has formulated a mid- to long-term AI strategy grounded in a sovereign AI approach. The strategy's core element is the implementation of the inclusive "AI for All" vision within the context of Korean ODA (K-ODA), with the objective of ensuring shared benefits for both South Korea and its partner countries. The approach of the Korea International Cooperation Agency (KOICA) is guided by four principles: the promotion of mutually beneficial artificial intelligence (AI) transformation, the contribution of a Korean-style AI model to the global community, the strengthening of AI capacity in developing countries, and the improvement of operational efficiency through the adoption of AI.

The Korea International Cooperation Agency (KOICA) is expanding its "AI-driven innovative ODA" initiatives across multiple sectors, including healthcare, smart cities, and digital public administration. A key focus of these initiatives is on collaboration with Global South partners (KOICA 2025). These initiatives seek to capitalize on South Korea's technological strengths, including semiconductors, cloud computing, and artificial intelligence, while aligning with the demographic trends, digital transitions, and data ecosystems of partner countries.

This transition is exemplified by the rapid growth and diversification of KOICA's AI project portfolio. There is a growing emphasis on health and governance, as well as an increasing reliance on public–private partnerships. Consequently, projects are evolving from pilot initiatives to long-term programs. Recent examples include the implementation of AI-based water resource management systems in Indonesia (2025, 3.1 billion KRW) and the introduction of AI-focused human resource development for industry in Bangladesh (2026, 39.3 billion KRW). In a broader context, KOICA has augmented the scale and scope of its AI-related ODA, encompassing digital infrastructure, governance, and AI workforce development (Kim and Chun 2025).

Ⅴ. South Korea's New ODA Initiative and "AI for the Global Community"

The OECD (2019) conceptualizes the evolution of government as a transition from analogue government to e-government, and further to digital government, where users actively interact with the state and co-produce public services. The advent of disruptive technologies, such as artificial intelligence and the Internet of Things, has precipitated a paradigm shift in the modus operandi of governments. These entities are undergoing a transformative process, becoming platform-based, networked systems that facilitate collaboration among citizens, businesses, and public institutions.

To this end, future AI for good governance initiatives should support leapfrogging in areas such as AI-augmented government, smart cities, digital twins, cloud-based data centers, AI-enabled public services, digital ID systems, and broader digital public infrastructure.

The "AI for All" initiative, as implemented in South Korea, is undergoing a systematic expansion to encompass the "AI for All Global Community." This evolution entails the utilization of artificial intelligence as a means to ensure inclusivity and to address global challenges, thereby contributing to the enhancement of society and the global community. It is imperative for South Korea to proactively identify areas where AI can be utilized to enhance the quality of government in developing countries, thereby facilitating the realization of fundamental principles of good governance. In accordance with the Fourth Comprehensive Basic Plan for International Development Cooperation, the Republic of Korea (hereafter South Korea) is obligated to augment its support for artificial intelligence (AI)-based digital government innovation in developing countries.

Conventional digital government projects have been insufficient to address the challenges faced by developing countries. To this end, future AI for good governance initiatives should support leapfrogging in areas such as AI-augmented government, smart cities, digital twins, cloud-based data centers, AI-enabled public services, digital ID systems, and broader digital public infrastructure. These initiatives will facilitate AI for good governance and eventually improve quality of life in developing countries. The design of these projects must align with the national vision and institutional readiness of each partner country, encompassing political commitment, ICT capacity, stakeholder coordination, data governance, cybersecurity, and accountability mechanisms.

It is imperative for South Korea to prioritize the promotion of participatory and accountable AI in the context of democracy and good governance in developing countries.

The effective design of projects necessitates a thorough evaluation of the partner country's preparedness, encompassing aspects such as political dedication, institutional congruence, stakeholder collaboration, ICT infrastructure, digital capability, citizen engagement, data governance, cybersecurity, and mechanisms for accountability. These elements are indispensable for ensuring the sustainability of AI-driven governance projects and their developmental ramifications.

The implementation of AI support should be firmly embedded in inclusive and secure digital public infrastructure, with a primary focus on high-impact areas such as anti-corruption, service delivery, and disaster management. Human oversight must remain central, and all systems should undergo impact assessments to address bias, privacy, and accountability risks as outlined in South Korea's AI legal framework.

It is imperative for South Korea to prioritize the promotion of participatory and accountable AI in the context of democracy and good governance in developing countries. The integration of AI into governance processes necessitates the engagement of citizens and civil society, complemented by mechanisms that ensure transparency, accountability, and the ability to address any resulting concerns. This comprehensive approach is essential for ensuring that AI contributes not only to administrative efficiency but also to inclusive, trustworthy, and resilient governance. It is imperative that AI initiatives for enhancing governance in developing countries prioritize governance as a primary concern, superseding the emphasis on technological aspects. In order to facilitate the alignment of K-ODA with global frameworks such as the Hiroshima AI Process, the UNESCO AI Ethics Recommendation, and the UN Digital Public Infrastructure, it is imperative for South Korea to prioritize the enhancement of institutional capacity, legal frameworks, and accountable, participatory public systems prior to—or in a concurrent manner with—the implementation of AI technologies.

References

Kim, Eunmee. 2021. "Directions for Development Cooperation in Governance in the Era of Digital Transformation (In Korean)." *Journal of International Development Cooperation* 16 (1): 101-118.

Joint Ministerial Press. 2025. "Action Plan for the Establishment of National AI Computing Center (SPC) (In Korean)." https://gonggam.korea.kr/newsContentView.es?code_cd=0114000000.

Kim, Woojin and Jeongwon Chun. 2025. "Exploring an Ethical Governance Framework for the Use of Artificial Intelligence (AI) in International Development Cooperation." *Journal of International Development Cooperation* 20 (1): 145-174. https://doi.org/10.34225/jidc.2025.20.1.145.

Korea International Cooperation Agency (KOICA). 2025. "Linking AI Transformation with International Development Cooperation." KOICA Press Release. October 29.

Korea International Cooperation Agency (KOICA). 2026. "KOICA Portal." https://www.koica.go.kr/koica_kr/901/subview.do.

Ministry of Foreign Affairs (MOFA). 2024. "Indo-Pacific Regional Meeting of the Second Summit for Democracy." https://www.mofa.go.kr/www/brd/m_26779/view.do?seq=537.

Ministry of Interior and Safety (MOIS). 2017. "Fifty Year History of Korea E-government" (In Korean). https://www.data.go.kr/data/15038598/fileData.do.

Moon, M. Jae. 2023. "Digital Government in Global Public Administration Issues (In Korean)." Unpublished Report.

Organization for Economic Co-operation and Development (OECD). 2025. "Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions." https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/governing-with-artificial-intelligence_398fa287/795de142-en.pdf.

Prime Minister's Secretariat. 2026. "The Fourth Comprehensive Basic Plan for International Development Cooperation (In Korean)." https://www.opm.go.kr/opm/news/press-release.do?mode=view&articleNo=161091.

Part 4

South Korea's Sovereign AI Initiative for Democracy - A Transferable Model for the Global Community

Jong Jin Lee
Senior Fellow, Institute for Peace and Unification Studies

Ⅰ. Introduction: Why AI Sovereignty matters for Democracy

The geopolitics of artificial intelligence has entered a decisive phase. As large language models (LLMs), autonomous agents, and AI-integrated defense systems proliferate, the question of who controls the core determinants of AI — data, compute, model weights, and alignment values — has become inseparable from questions of political sovereignty and democratic legitimacy. For decades, liberal democracies outsourced digital infrastructure to a handful of Silicon Valley platform companies. The costs are now visible: algorithmic radicalization, epistemic fragmentation, and "epistemological colonization," in which foreign-designed AI systems quietly impose particular worldviews on national publics without democratic consent.

South Korea's response is distinctive. Rather than importing AI infrastructure or racing for commercial dominance, Korea has pursued a third path: Sovereign AI as a democratic institution — embedding democratic values

into AI training, establishing public compute infrastructure, securing rights over model weights, and institutionalizing civic oversight.

Korea has codified these commitments in the AI Basic Act (2024/2026) — one of the most comprehensive national AI governance frameworks in existence. The result is a model that is globally transferable, particularly to middle-power democracies seeking an alternative to both Silicon Valley techno-liberalism and Chinese state-AI authoritarianism.

Ⅱ. The Democratic Stakes: AI and the Crisis of Political Participation

Before examining Korea's model, it is essential to understand the democratic threat that Sovereign AI is designed to address. Scholars have increasingly documented a structural tension between AI-optimized systems and democratic deliberation.

Song Kyungho (2025) identifies a structural shift in post-AI governance — "the technologization of politics" — in which AI reconfigures politics as a problem of computational optimization. This manifests through three dangerous archetypes: the Necromancy Model (AI simulating political figures to fragment discourse); the Utilitarian Machine (AI policy engines maximizing efficiency at the expense of democratic accountability); and the Philosopher King Model (AGI-based governance that renders human political agency structurally redundant).

Each archetype instantiates a "participation paradox": technology ostensibly expands civic engagement while actually displacing genuine political agency. Citizens become data points; deliberation becomes optimization; accountability dissolves into algorithmic black boxes.

This structural threat is compounded by the ideological projects of dominant technology entrepreneurs. Peter Thiel's libertarian exit from democratic accountability, Elon Musk's Department of Governement Efficiency (DOGE) model of technocratic state management, and Alexander Karp's technology republic vision all converge on one conclusion: democratic

deliberation is an inefficiency to be engineered away. Farrell and Newman's diagnosis of "digital feudalism" captures the result — private AI infrastructure functioning as sovereignty without democratic legitimacy (Farrell and Newman, 2023).

For South Korea, a frontline democracy geographically adjacent to North Korea's AI-enabled influence operations and geopolitically entangled in US-China technology competition, these threats are not abstract. They are existential. The Korean model of Sovereign AI is therefore not merely a policy preference but a strategic necessity.

Ⅲ. South Korea's Sovereign AI Framework: Architecture and Institutions

Korea's AI governance framework has evolved through three distinct phases since 2016, each layering new institutional capacity onto the previous foundation.

The Korean model of Sovereign AI is therefore not merely a policy preference but a strategic necessity.

1. Phase One: Ethical Foundation (2016–2020)

The 2016 AlphaGo shock catalyzed Korea's AI governance consciousness. In response, the government established the AI Ethics Research Committee, which produced the Seoul PACT — Korea's first national AI ethics guidelines — in March 2018. The 2020 National Guidelines for AI Ethics elaborated this into a multilayered framework anchored by three core principles:

Table 1. Three Core Principles and Key Requirements of Korea's National Guidelines for AI Ethics (2020)

Principle	Key Requirements
Respect for Human Dignity	Safeguarding human rights, protection of privacy, respect for diversity, prevention of harm
Common Good of Society	Public good, solidarity, accessibility for the vulnerable, equitable distribution of AI benefits
Proper Use of Technology	Data management, accountability, safety, transparency, explainability

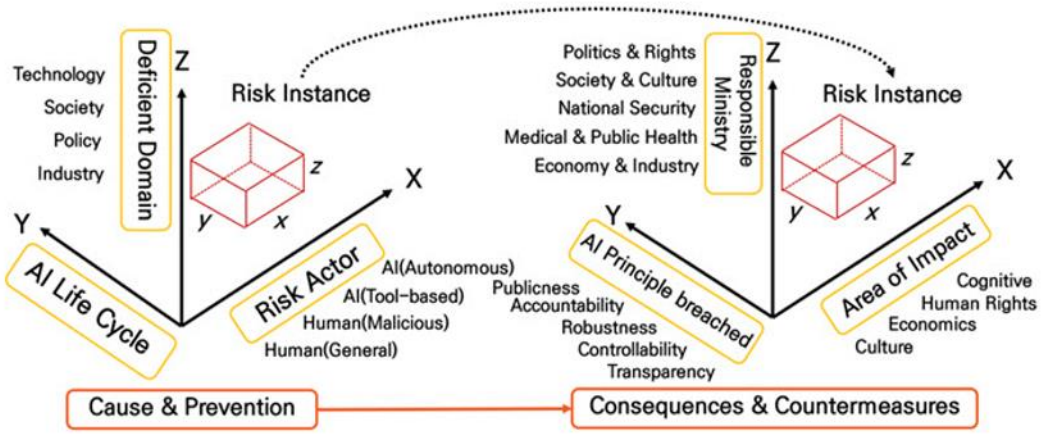
2. Phase Two: Institutional Architecture (2020–2025)

Building on this ethical foundation, Korea constructed a layered institutional architecture. The AI Basic Act, passed in December 2024 and entering full effect in January 2026, is the centerpiece, with three key institutional innovations.

First is the National AI Strategy Committee. Chaired by the President and composed of cabinet ministers and civilian experts, it serves as both advisory and decision-making authority for AI policy. Following the 2025 change of government, the Committee's authority was strengthened as AI competitiveness became the highest national policy priority.

Second is the Korea AI Safety Institute (AISI). Established in November 2024, Korea became the sixth country globally to create a national AI Safety Institute. Korea AISI serves as the national hub for AI safety, with seven statutory functions: risk definition and analysis, safety policy research, evaluation criteria development, safety technology standardization, international cooperation, frontier AI safety assurance, and other presidential decree functions.

Figure 1. Risk-Mapping Methodology of Korea AISI by Kim (2026)



As seen in the Figure 1, AISI's distinctive contribution is its risk-mapping methodology — a three-dimensional analytical framework (MECI: mutually exclusive, collectively inclusive) that maps risks along axes of risk actor, AI life cycle stage, and deficient domain, then assigns countermeasures to responsible ministries. This moves beyond conventional risk cataloguing toward dynamic, actionable risk governance.

Third is the AI Policy Center. It is tasked with developing comprehensive AI policies and, critically, promoting the establishment and dissemination of international norms. This function explicitly positions Korea as a norm exporter rather than merely a norm taker in global AI governance.

3. Phase Three: The K-AI Initiative and Sovereign Model Development

The K-AI Initiative represents Korea's most ambitious Sovereign AI project: the development of a globally competitive LLM that embeds Korean values and democratic processes directly into its technical core — specifically its weights. The initiative was launched with fifteen corporate consortia applicants, from which five teams were selected in August 2025 — Naver Cloud, Upstage, SK Telecom, NC AI, and LG AI Research. Following a first-stage evaluation

in January 2026 that eliminated Naver Cloud and NC AI, three remaining consortia (LG AI Research, SK Telecom, and Upstage) are currently advancing through the second phase of testing, with a final evaluation scheduled for December 2026 and two finalists to be designated by 2027. Throughout this process, Korea AISI conducts safety and trustworthiness evaluations alongside performance assessments.

The theoretical framework (proposed framework not yet implemented policy) underlying the K-AI Initiative rests on five pillars of democratic AI architecture:

Table 2. Five-Layer Democratic AI Architecture of the K-AI Initiative (Proposed Framework)

Layer	Component	Democratic Rationale
Data	Democratic Value Alignment Datasets	Establishes a foundational worldview grounded in Korean values and norms, preventing epistemic colonization by foreign corpora
Infrastructure	Public GPU Farms	Dismantles technology monopolies; universal access for academia and SMEs; sovereignty over compute
Weights	Auditable Weight Control	Secures modification and audit rights over the core determinants of AI values; "cognitive shield" against algorithmic bias
Learning	Citizen Jury RLHF [Aspirational]	Replaces closed Silicon Valley labeler decisions with deliberative democratic input; reward functions based on social consensus
Algorithms	Pluralistic AI Ecosystem	Guards against single-model monopoly; enables NGO, academic, and SME models to coexist and check filter bubbles

Ⅳ. Democratic Alignment: Korea's Distinctive Contribution

The most innovative element of Korea's proposed democratic AI architecture is a mechanism to replace closed-door labeling decisions — currently made by a handful of contractors in AI companies — with deliberative democratic processes.

South Korea's AI policy can be assessed from the lens of "Democratic Alignment." This concept captures a genuine orientation in Korea's AI policy: the question of who decides what values are embedded in AI systems, and through what processes. While most national AI strategies focus on safety or competitiveness, Korea's National Guidelines for AI Ethics (2020) and AI Basic Act (2026) both position democratic legitimacy and civic participation as foundational governance concerns. This section presents that orientation's normative elaboration, distinguishing codified provisions from proposals or aspirations.

1. Democratic RLHF: Citizen Juries as Alignment Mechanism[1]

The most innovative element of Korea's proposed democratic AI architecture is a mechanism to replace closed-door labeling decisions — currently made by a handful of contractors in AI companies — with deliberative democratic processes. "Democratic Reinforcement Learning from Human Feedback (RLHF)" involves randomly selected citizen juries whose deliberated judgments are reflected in AI reward functions, addressing a fundamental democratic deficit: the values embedded in frontier AI models are determined by private actors without democratic accountability. Korea's proposed framework would democratize this process. The AI Basic Act's Article 27

[1] "Democratic RLHF" and the Citizen Jury mechanism described in this section are the author's policy proposals, not provisions of current Korean law. No official government document — including the AI Basic Act, the National AI Ethics Guidelines, or the K-AI Initiative framework — specifies citizen juries as a component of AI training. The proposals are presented here as a normative extension of Korea's existing civic participation orientation, grounded in international deliberative democracy precedents.

provides a partial legislative foundation, requiring MSIT to collect opinions from diverse sectors of society in establishing ethical principles — but this falls well short of the binding citizen jury mechanism proposed here.

Under this framework, AI-generated decisions in the public sector must be subject to human deliberation and critical review.

Precedents exist for the broader approach. Taiwan's pol.is platform engaged nearly 12 million citizens in AI-assisted deliberation (Tang, 2025), with AI serving as what Google DeepMind researchers call a "deliberative mediator" rather than a decision-maker. The Habermas Machine experiment demonstrated that AI-mediated deliberation can produce consensus statements rated higher in clarity, fairness, and representativeness than those produced by human facilitators. Korea itself has piloted deliberative polling through the National Assembly (2023), providing a domestic institutional basis for scaling such mechanisms to AI governance.

2. Procedural Delay as Democratic Safeguard[2]

Against the efficiency logic of Big Tech — which treats deliberative friction as overhead to be engineered away — this analysis argues for "procedural delay" as a structural feature of democratic AI governance, building on the human oversight requirements already embedded in the AI Basic Act. Under this framework, AI-generated decisions in the public sector must be subject to human deliberation and critical review.

This principle applies particularly to "epistemic" domains: AI interpretation of national history, security assessments, judicial decisions, and policy recommendations. The argument is not that AI is wrong in these domains, but that democratic legitimacy requires human deliberation

[2] "Procedural delay" as a named principle does not appear in Korean official documents. However, its underlying rationale — that AI-generated outputs in high-stakes public domains must be subject to mandatory human review — is partially codified in the AI Basic Act. Article 32 requires providers of high-impact AI to establish and operate human management and supervision mechanisms and to explain AI-derived results to affected parties. This statutory human oversight requirement provides a legal anchor for the broader "procedural delay" concept articulated here.

regardless of AI accuracy. In this light, the human-in-the-loop oversight mandated by Article 32 of Korea's AI Basic Act should not be treated as a bureaucratic hurdle, but as an institutional mechanism that operationalizes "procedural delay" by ensuring high-stakes automated decisions are legally paused for deliberative human judgment.

Of the three components, civic oversight has the strongest basis in enacted law. The AI Basic Act (Articles 27–29) establishes a multi-layered oversight structure that this analysis extends into three operational layers. Items marked [Codified] reflect provisions already in force; items marked [Proposed] reflect the author's recommended extensions beyond current law:

- Offensive Red Teams [Codified — Article 29, AISI mandate]: Systematic adversarial testing of AI models for biases and potential authoritarian misuse, with results made publicly available. Korea AISI is already operationalizing this through its dual-use CBRN-E task force and deepfake detection research. The extension proposed here is that red-team findings be made available not only to government ministries but to the general public.

- Self-Regulatory Ethics Committees [Codified — Article 28]: The AI Basic Act authorizes companies, universities, and research institutes to form Private Autonomous AI Ethics Committees with authority to verify compliance, investigate human rights concerns, and conduct ethics education. These committees must include external members capable of evaluating ethical and social validity, and must not consist solely of one gender. Proposed extension: grant independent civil society organizations — not only institutional actors — standing to conduct equivalent oversight, and make findings subject to regulatory follow-up.

- Auditable Sandboxes [Proposed — not yet codified]: External auditors with access rights to internal model structures and training data to verify technical malfunctions or human rights violations, balanced against trade secret protection. The AI Basic Act (Article 40) grants MSIT authority to demand data and inspect AI business operators, but this is a government power rather than a civil society right. The "Auditable Sandbox" concept proposed here would extend equivalent access — under appropriate confidentiality protections — to independent researchers and civil society organizations, bringing Korean practice closer to the EU AI Act's provisions for third-party auditing of high-risk systems.

Taken together, these three layers — one fully codified, one partially codified, one proposed — represent a trajectory rather than a completed architecture. Korea's AI Basic Act has established the institutional foundations for civic participation in AI governance; the proposals here argue for deepening those foundations toward a model in which citizens are active agents of AI accountability, not merely protected from AI harms. The distance between Korea's current provisions and the full democratic alignment framework is a measurable legislative challenge — specific and addressable through concrete policy action.

Ⅴ. The International Dimension: Korea as Norm Exporter

The distance between Korea's current provisions and the full democratic alignment framework is a measurable legislative challenge — specific and addressable through concrete policy action.

Korea's Sovereign AI framework has explicit international ambitions. The AI Basic Act establishes the AI Policy Center with a mandate to "promote the establishment and dissemination of international norms." Korea AISI participates in the international AISI

Network (10 countries) and has signed MOUs with France, the US, Poland, Singapore, and ASEAN partners — reflecting a strategic positioning of Korea as a middle-power norm entrepreneur in global AI governance.

1. The Middle Power Advantage

Korea's middle-power status is a governance asset in the current AI landscape. Unlike the US or China, Korea does not trigger hegemonic anxieties when proposing international AI standards, while possessing distinctive asymmetric technological credibility: world-leading hardware capacity paired with a rapidly advancing but not yet frontier-level software stack.

On hardware, Korea's position is unambiguous. Samsung and SK Hynix together dominate global high-bandwidth memory (HBM) production — the critical component powering AI accelerators worldwide — with SK Hynix holding approximately 57% of the HBM market as of 2026 (CNBC 2026.4). Both are confirmed suppliers for Nvidia's next-generation Vera Rubin architecture, and Korea's DRAM and advanced packaging capabilities are widely regarded as irreplaceable in the near term.

Unlike the US or China, Korea does not trigger hegemonic anxieties when proposing international AI standards, while possessing distinctive asymmetric technological credibility: world-leading hardware capacity paired with a rapidly advancing but not yet frontier-level software stack.

On software, the picture is more nuanced. Korean LLMs — including LG's EXAONE, Naver's HyperClova X, and Upstage's Solar series — have demonstrated competitive performance on multilingual benchmarks, and the K-AI Initiative targets 95% of frontier model performance. However, these models currently operate well below the scale of leading US frontier systems, and domestic GPU infrastructure remains substantially dependent on imported Nvidia hardware. Korea's LLM capabilities are best characterized as rapidly advancing rather than frontier-equivalent — a distinction that honest governance advocacy must acknowledge.

This asymmetry is a source of credibility, not weakness. Korea is not seeking to dominate global AI; it is seeking to govern it responsibly while building indigenous capacity. The result is a form of "techno-diplomatic" influence that operates differently from traditional power politics: Korea can convene conversations that neither Washington nor Beijing can host without triggering strategic suspicion — particularly valuable in the Indo-Pacific, where partners like Japan, Australia, and ASEAN members seek governance frameworks that protect technological sovereignty without forcing sides in US-China competition.

Korea is not seeking to dominate global AI; it is seeking to govern it responsibly while building indigenous capacity.

This positioning is no longer merely aspirational. In May 2026, the Korean government formalized its role as a multilateral AI convener by launching the Global AI Hub — a joint initiative signed with nine major UN agencies (ILO, IOM, ITU, UNDP, UNEP, UNHCR, UNICEF, WFP, WHO) and five multilateral development banks (WB, ADB, IDB, EBRD, CABEI). Operating under the vision of "AI for All, AI to Solve Global Challenges," the Hub is designed to integrate fragmented international AI capacities into a shared infrastructure platform, enabling coordinated responses to climate change, public health, food security, forced displacement, and labor transitions. Critically, the Hub's Joint Statement — signed in Seoul on 21 May 2026 — explicitly positions the initiative not as a Korean-directed program but as a collaborative capability co-shaped by the Participating Organizations, with additional UN entities invited to join over time. This architecture embodies precisely the "horizontal partnership" logic that distinguishes Korea's international AI strategy: Korea as convener and funder, not as norm hegemon.

2. Three Transferable Frameworks

Unlike technology transfer models in which a leading state exports pre-packaged governance solutions, Korea's approach is premised on genuine co-production — shared institutional design, mutual accountability, and reciprocal learning across partners at varying stages of AI development.

Korea's model offers three distinct frameworks for partner democracies. The first — AI Security — provides concrete mechanisms for defending against "epistemological colonization": the "Cognitive Shield" framework secures sovereign control over model weights, establishes national corpora reflecting domestic values, and builds red-team capacity to identify AI-enabled influence operations. The second — AI Ethics — constitutes a comprehensive democratic alignment architecture: Korea's ten-requirement ethics framework, "Democratic RLHF" model, and procedural delay principles, explicitly designed to distinguish engineering-based safety from normative governance. The third — Technical Standards — advances international "Auditable Sandbox" standards, Hub-and-Spoke interoperability norms, and open-source foundation model development with EU and Japan partners, designed to prevent single-source monopolies while enabling sector-specific AI development across academic, NGO, and SME communities.

3. Strategic Alliances: Beyond Donor-Recipient Dynamics

A distinctive feature of Korea's international AI strategy is its emphasis on "horizontal and strategic partnerships that transcend the traditional donor-recipient dynamic." Unlike technology transfer models in which a leading state exports pre-packaged governance solutions, Korea's approach is premised on genuine co-production — shared institutional design, mutual accountability, and reciprocal learning across partners at varying stages of AI development.

Korea's role is most credibly that of a norm entrepreneur and convening partner, not a norm hegemon.

This orientation is already visible in Korea's multilateral engagements. At the research level, Korea co-participates in international AISI tracks on deepfakes, multicultural AI testing, and risk identification — contributing as a peer, not a recipient. At the policy level, Korea's participation in the Pax Silica Initiative alongside Australia, Japan, and the UK, and its EU-ROK joint semiconductor research under the Chips Joint Undertaking, demonstrate multilateral engagement on AI supply chain security and semiconductor resilience. At the standards level, Korea is co-developing open-source foundation model components with EU and Japanese partners, and sharing its AI safety evaluation frameworks with ASEAN members.

The logic of horizontal partnership also applies to the governance frameworks in this report. The three transferable frameworks — AI security resilience, democratic alignment, and pluralistic technical standards — are most likely to gain international traction when co-developed with partner democracies rather than exported as a finished Korean model. Korea's role is most credibly that of a norm entrepreneur and convening partner, not a norm hegemon. The May 2026 launch of the Global AI Hub — a Korea-hosted, UN-co-shaped platform uniting nine international organizations and five multilateral development banks under the vision of "AI for All, AI to Solve Global Challenges" — stands as the most institutionally concrete demonstration of this horizontal partnership logic to date.

Translating this into concrete action requires three priorities. First, Korea should accelerate the AISI International Network's expansion — particularly toward middle-power democracies in Africa, Latin America, and South Asia — positioning Korea as a co-leader in global AI safety capacity-building alongside, not subordinate to, the US-UK-EU axis. Second, Korea should work actively within UNESCO, OECD, and UN frameworks to establish its AI Basic Act and AISI as reference models for member states, building on the precedent set by the EU AI Act's international influence. Third, Korea should commission joint research with Japan and Australia on AI supply

chain security, hardware trust frameworks, and model integrity standards — an area where the three countries′ combined expertise constitutes a uniquely credible Indo-Pacific contribution to global AI norm-setting.

The ultimate measure of Korea's international AI strategy will not be the volume of MOUs signed or framework documents published. It will be whether the governance architecture described in this report — democratic alignment, civic oversight, cooperative sovereignty — becomes a reference point that other middle-power democracies adapt and build upon.

VI. Transferability: Conditions and Constraints

While Korea's model offers genuine transferable value, adoption requires contextual adaptation. Three conditions are necessary for successful transfer:

1. Institutional Preconditions

The ultimate measure of Korea's international AI strategy will not be the volume of MOUs signed or framework documents published.

Korea's model rests on foundations that took decades to build: a mature civil society capable of meaningful AI oversight; strong rule of law capable of enforcing auditable sandbox provisions; and technical capacity in both the public sector (Korea AISI) and private sector (Samsung, SK Hynix, Kakao, Naver). Democracies lacking these foundations will need to invest in institutional capacity before attempting to replicate the full framework.

2. Political Preconditions

Korea's AI governance has maintained bipartisan continuity across multiple government changes, reflecting an underlying consensus that AI governance is a national security issue, not a partisan one. Democracies where AI governance

is politically contested — where opposition parties might systematically undermine AI safety regulations for competitive advantage — will face structural obstacles to implementing Korea's model.

3. Modular Adoption

For democracies that cannot immediately implement Korea's full framework, a modular adoption approach is recommended. Priority modules include:

- Immediate (Year 1–2): Adopt Korea's ten-requirement AI ethics framework; establish a national AI Safety Institute with mandate for international cooperation; enact basic auditable sandbox provisions.

- Medium-term (Year 2–4): Develop national AI corpora with democratic value alignment; establish Public GPU access programs for academia and SMEs; build civil society AI auditing capacity.

- Long-term (Year 4+): Develop sovereign foundation models with Democratic RLHF mechanisms; establish international interoperability standards with partner democracies; build Hub-and-Spoke pluralistic AI ecosystems.

Ⅶ. Conclusion

The central question of AI governance in the coming decade is not merely technical — it is political. It concerns who controls the values embedded in AI systems that increasingly mediate political discourse, public administration, security decisions, and cultural identity. The answer will determine whether AI becomes an instrument of democratic self-determination or an engine of authoritarian control.

South Korea's Sovereign AI Initiative for Democracy offers a compelling answer: AI systems should be democratically aligned, civically governed, technically sovereign, and internationally transferable. This is not utopianism — it is a concrete institutional architecture that Korea has been building for nearly a decade, and which is now codified in the AI Basic Act.

The model is not perfect. Korea's democratic AI aspirations outrun current technical capacities in some domains; civil society oversight mechanisms are still nascent; and the tension between commercial AI competitiveness and democratic alignment remains unresolved. But these are engineering and institutional challenges — solvable. The conceptual framework is sound, the institutional architecture is established, and the political will is present. For the global democratic community navigating an AI landscape increasingly dominated by authoritarian state AI and unaccountable private AI, Korea's model represents a democratic third way that makes it not just a Korean achievement, but a global public good.

South Korea's Sovereign AI Initiative for Democracy offers a compelling answer: AI systems should be democratically aligned, civically governed, technically sovereign, and internationally transferable.

References

Bae, Hackyoung. 2026. *Sovereign AI and ROK-U.S. Cooperation*. Seattle: The National Bureau of Asian Research.

Draudt-Véjares, Darcie and Seungjoo Lee. 2026. "Governing AI in the Shadow of Giants: Korea's Strategic Response to Great Power AI Competition." Washington, D.C.: Carnegie Endowment for International Peace.

Center for Security and Emerging Technology (CSET). 2025. "Framework Act on the Development of Artificial Intelligence and Establishment of Trust (English translation)." Washington, D.C.: Georgetown University.

Coeckelbergh, Mark. 2024. *Why AI Undermines Democracy and What to Do About It*. Hoboken: Wiley.

CNBC. 2026. "SK Hynix Posts Record First-Quarter Profit as Memory Prices Climb." April 23. https://www.cnbc.com/2026/04/23/sk-hynix-posts-record-first-quarter-profit-as-memory-prices-climb.html. (Accessed: April 23, 2026)

Farrell, Henry and Abraham L. Newman. 2023. *Underground Empire: How America Weaponized the World Economy*. New York: Henry Holt.

Karp, Alexander C. and Nicholas W. Zamiska. 2025. *The Technological Republic*. New York: Penguin Random House.

Kim, Myung-Joo. 2026. "South Korea's Approach to AI Governance: Mitigating Risks and Ensuring AI Safety." Seattle: The National Bureau of Asian Research.

Korea Herald. 2025. "LG, SKT, Naver Among Five Selected for Korea's Sovereign AI Push." August 4. https://www.koreaherald.com/article/10337170. (Accessed: August 4, 2025)

Korea Herald. 2026. "Korea's Sovereign AI Push Stumbles over 'From Scratch' Rule." January 20. https://www.koreaherald.com/article/10590538. (Accessed: February 20, 2026)

Landemore, Hélène. 2023. "Fostering More Inclusive Democracy with AI." *Finance & Development* 60, 4: https://www.imf.org/en/Publications/fandd/issues/2023/12/fostering-more-inclusive-democracy-with-AI-landemore. (Accessed: January 24, 2026)

Ministry of Science and ICT (Republic of Korea). 2020. "The National Guidelines for Artificial Intelligence (AI) Ethics." Seoul: KISDI.

Ministry of Science and ICT (Republic of Korea). 2024. "AI Basic Act." Seoul: National Assembly of Korea.

Schaake, Marietje. 2024. *The Tech Coup: How to Save Democracy from Silicon Valley*. Princeton: Princeton University Press.

Song, Kyungho. 2025. "Democracy After AI: Technological Possibilities and the Paradox of Participation." *Legislation and Policy* 17, 2: 39–72.

Tang, Audrey. 2025. "Is Digital Collective Intelligence Still the Future of Democracy?" Presented at the 4th Hankyoreh People and Digital Forum. Seoul, Republic of Korea.

Tessler, M. H., Michiel A. Bakker, Daniel Jarrett, Hannah Sheahan, Martin J. Chadwick, Raphael Koster, Georgina Evans, Lucy Campbell-Gillingham, Tantum Collins, David Sumers, Edward Hughes, Anton Novikov, Kris Cao, Ilia Shumailov, Silvia Chiappa, Matthew Botvinick, and Christopher Summerfield. 2024. "AI Can Help Humans Find Common Ground in Democratic Deliberation." *Science* 386, 6719. https://doi.org/10.1126/science.adh2889.

Thiel, Peter. 2009. "The Education of a Libertarian." *Cato Unbound*. April 13. https://www.cato-unbound.org/2009/04/13/peter-thiel/education-libertarian/. (Accessed: March 13, 2026)

Asia Democracy Research Network

South Korea's Approach to **Democratic AI Governance**

Asia Democracy Research Network (ADRN) is an independent network of democracy research institutions across Asia.
It analyzes challenges and tasks facing democracy in the region and expands the scope of action-oriented, policy-driven research that supports the advocacy activities of Asian civil society organizations in promoting, consolidating, and deepening democracy.

ADRN Secretariat
East Asia Institute
1 Sajik-ro 7-gil, Jongno-gu, Seoul 03028, Republic of Korea