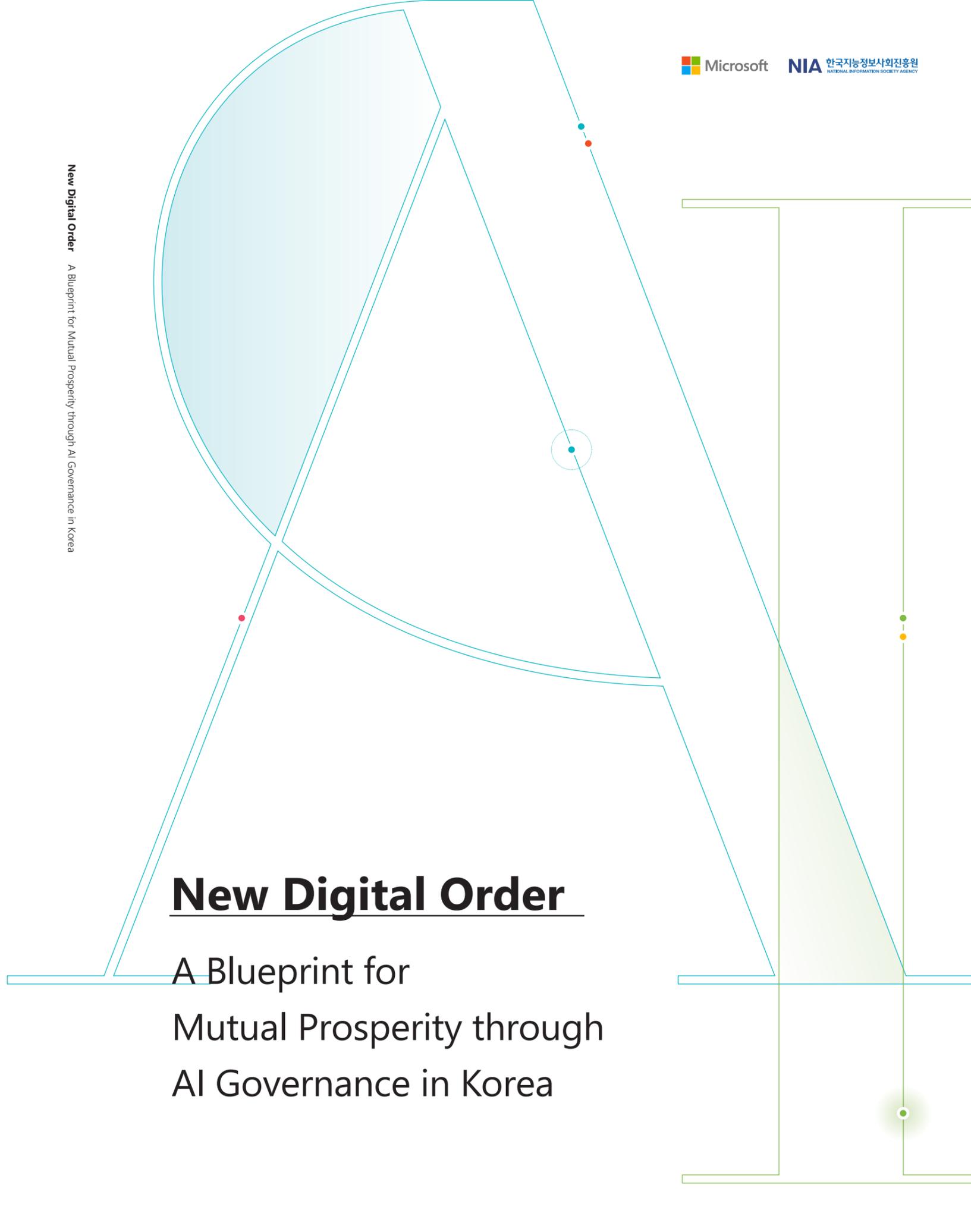


New Digital Order A Blueprint for Mutual Prosperity through AI Governance in Korea

New Digital Order

A Blueprint for
Mutual Prosperity through
AI Governance in Korea



Acknowledgments

The publication of this positioning paper was supported by the National Information Society Agency (NIA).

This paper also benefited from the support of Dr. Moon Jung Wook from the Korea Information Society Development Institute (KISDI), Lee Kang-hae from the Telecommunications Technology Association (TTA), and Prof. Lim Yong of the Seoul National University AI Policy Initiative.

New Digital Order

A Blueprint for
Mutual Prosperity through
AI Governance in Korea

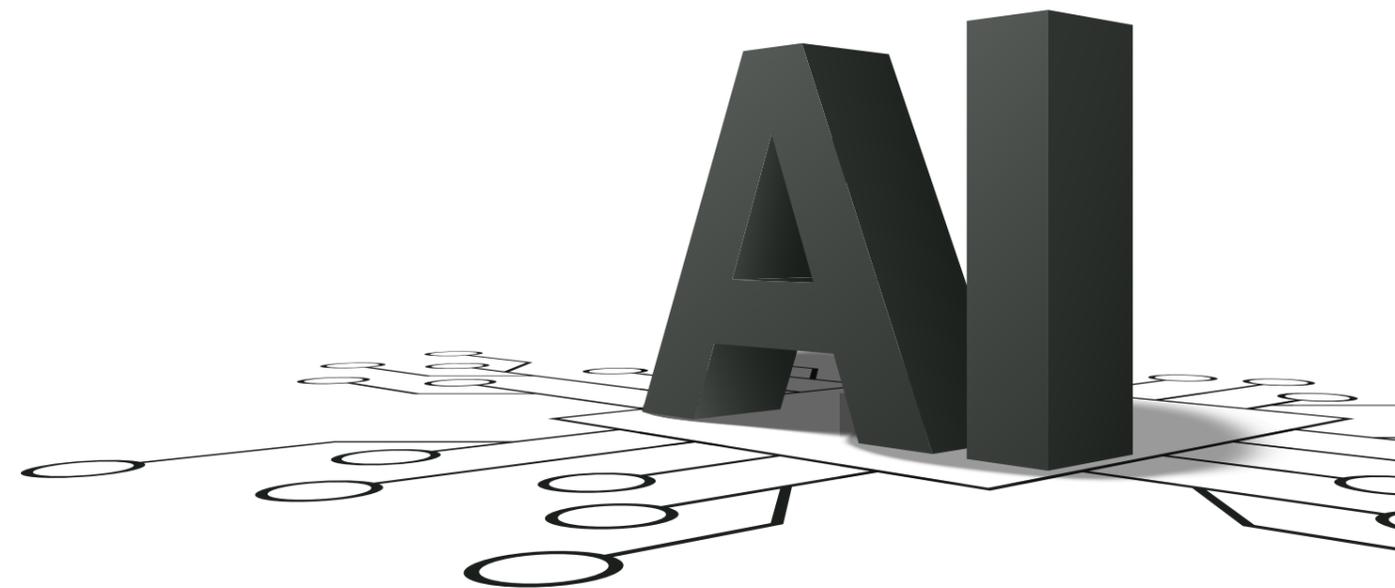


Table of contents



• Foreword 006

• Part 1 012

The development and utilization of Artificial Intelligence (AI) through responsible AI governance is a crucial element of a digital society of mutual prosperity

- 11 The AI industry has emerged as a core industry for Korea to achieve the goal of coexistence and mutual prosperity.
- 12 Korea's Digital Bill of Rights is an apt starting point for discussing the establishment of secure and trustworthy AI governance.

• Part 2 019

Korea, the international society, and the tech industry are on an organic journey towards a digital society of mutual prosperity

- 16 Korea's principle-based approach to AI governance for the goal of a digital society of mutual prosperity is in line with the response of the international community and the tech industry, which argues for the development and use of AI under the three themes of responsible AI, inclusive AI, and sustainable AI.
- 17 Korea's approach to responsible AI
 - Principle 3: Creation of a safe and trustworthy digital society
 - Principle 4: Promotion of autonomous and creative digital innovation
- 31 Korea's approach to inclusive AI
 - Principle 1: Protection of digital freedom and rights in the digital
 - Principle 2: Guarantee of fair access and equitable opportunities in the digital
- 42 Korea's approach to sustainable AI
 - Principle 5: Promotion of the well-being of all humankind



• Conclusion 052

• Appendices 054

- 48 Appendix A Current status of AI governance in Korea
- 56 Appendix B Recent global developments in AI governance
- 64 Appendix C Mapping of global AI governance initiatives against Korea's Digital Bill of Rights
- 70 Appendix D Best practices from Microsoft to build safe and trustworthy AI systems for a digital society of mutual prosperity
- 83 Appendix E Case studies on the potential impact of the AI industry on the Korean economy

• References 096



Foreword

Willy Cho,
Area Vice President,
Microsoft Korea



Korea's AI Transformation

At the forefront of a technological revolution, Korea's journey with artificial intelligence (AI) epitomizes a harmonious blend of innovation and responsibility. As we embark on a new chapter in human history, our approach to AI is not merely about technological advancements but about sculpting a future that resonates with our societal ethos and global ambitions.

South Korea is on a fast track to achieve AI Everywhere, with instances of AI innovation manifesting across all facets of society. Key industries in Korea are at a transformative inflection point, rapidly embracing AI technologies, and Microsoft has been an indispensable partner in this journey.

Example 1) An exemplary case of this collaboration is in the field of senior healthcare, where we are addressing Korea's challenge of an aging population. Together with Microsoft, Professor Howard Lee from Seoul National University's Center for Convergence Approaches in Drug Development is conducting research on algorithms aimed at enhancing drug repurposing, which involves discovering new therapeutic uses for drugs, as well as AI-driven improvements to clinical trial design. The outcomes of this research include valuable insights into maximizing cost-effectiveness and improving the accuracy of clinical trials.

Example 2) Microsoft is also supporting research on deep learning for structure-based drug design, led by Professor Choi Sun at Ewha Woman's University's Global AI Drug Discovery Research Center. This research has succeeded in rapidly screening molecules with targeted properties and compounds with high binding affinity by analyzing protein and compound data with AI. Microsoft's domestic AI research collaborations are contributing to significantly reducing the time and cost of drug development in healthcare and improving drugs' medical efficacy, thereby easing the financial burden of managing geriatric diseases on the national healthcare system.

Example 3) More recently in March 2024, the potential of AI in educational settings has been explored through the lens of a project at the Korea Advanced Institute of Science and Technology (KAIST). Professor Alice Oh and her team embarked on a journey to harness the educational benefits of AI while mitigating the risk of it being used for shortcuts in academic work. Their solution came in the form of a chatbot developed with support from Microsoft Research's Advancing Foundation Models Research (AFMR) initiative, launched in April 2023. This initiative aims to propel the development and application of foundation models across various disciplines by providing academic researchers with access to advanced AI models through Azure AI Services. The chatbot developed by Oh and her team is designed to assist English as a Foreign Language (EFL) students with their essay writing, offering guidance without writing the essays for them. Throughout a semester, 213 EFL students interacted with the chatbot, engaging with

it as an intelligent peer and utilizing its feedback to refine their essays. The success of the project underscores the significant potential of generative AI in education, particularly in fostering a deeper understanding of problem-solving processes among students.

The impact of AI extends beyond the healthcare sector, permeating major domestic industries such as semiconductors, education, batteries, finance, telecommunications, manufacturing, and content.

As AI continues to drive scientific innovation and socioeconomic transformation, Korea is gearing up to reap its benefits to the fullest. It is also important to note that the foundation of this transformation lies in the hyperscale cloud infrastructure. This infrastructure provides the scalability and flexibility required for rapid innovation and robust AI applications.

AI Governance in Korea

In pace with the rapid technological advancements and transformation triggered by AI, Korea was one of the first countries in the world to release a set of widely applicable principles that highlight the importance of grounding and guiding the use and development of technology such as AI in the digital space. 'The Digital Bill of Rights: Charter for the Values and Principles for a Digital Society of Mutual Prosperity' lays out the guiding framework for Korea's vision on AI and influences the country's legislative, regulatory, and operational approaches to AI. Korea's vision for AI, as articulated by its government, is rooted in three core principles: 1) Responsible AI, 2) Inclusive AI, and 3) Sustainable AI.

Microsoft's own principles for AI closely align with the principles of the 'Digital Bill of Rights.' Our focus is on Responsible AI, which emphasizes transparency, fairness, ethics, and accountability, while considering societal impacts and privacy. We also aim for Inclusive AI, ensuring equitable access and benefits for everyone. Moreover, we are dedicated to Sustainable AI, utilizing it for the betterment of humanity and minimizing adverse effects.

Moreover, our commitment is also to empowering a vibrant and open market for AI to flourish. During the Mobile World Congress (MWC) in March this year, we unveiled our commitment to responsible AI through the announcement of our 'AI Access Principles.' The 11 principles fall under 3 major themes: 1) Providing access and support for AI developers who create models and applications, 2) Ensuring choice and fairness across the AI economy, and 3) Meeting our societal responsibilities.

Microsoft AI Access Principles

Provide access and support for AI developers who create models and applications.

- 1 As we grow chip capacity, we are expanding Microsoft's cloud computing AI infrastructure to enable the training and deployment of more foundation models, both proprietary and open source.
- 2 We are making AI models and development tools broadly available to software applications developers around the world, so every nation can develop its own AI economy.

Ensure choice and fairness across the AI economy.

- 3 We are making available public APIs to enable developers to access and use AI models we host.
- 4 We are supporting a common public API to enable network operators to support software developers.
- 5 Developers may choose how to distribute and sell their AI models, tools and applications for deployment and use on Microsoft Azure, whether via the Azure Marketplace or directly to customers.
- 6 We respect the needs of developers by ensuring we do not use any non-public information or data from the training, building, deployment, or use of developers' AI models in Microsoft Azure to compete with those models.
- 7 We enable customers using Microsoft Azure to switch to another cloud provider by easily enabling them to export and transfer their data.

Meet our societal responsibilities.

- 8 We are supporting the physical and cyber security needs of all the AI models and applications that run in our AI datacenters.
- 9 We are applying a strong Responsible AI Standard to keep people at the center of AI design decisions and respect enduring values like fairness, reliability, safety, privacy, inclusiveness, transparency, and accountability.
- 10 We are investing in initiatives to spread AI skilling broadly around the world.
- 11 We are managing our AI datacenters in an environmentally sensitive manner and using AI to advance environmental sustainability needs.

The 'AI Access Principles' serve as our compass, guiding Microsoft's role and responsibility as an AI innovator and market leader. Our principles signal a profound shift, they pledge Microsoft to unprecedented investments, robust business partnerships, and expansive programs aimed at fostering innovation and competition. We recognize that AI's transformative potential extends beyond corporate boundaries; it touches lives globally. By articulating these principles, we commit to providing broad technology access empowering organizations and individuals worldwide to wield AI for the greater good. Our initiatives around the world, including substantial AI datacenter investments and skilling programs, underscore our dedication to translating these principles into action.

Road to 2024 Seoul AI Safety Summit

As we approach the 2024 Seoul Safety Summit in May, we reflect on the strides made in the field of AI and the importance of ensuring its responsible development and deployment. The Summit serves as a platform for thought leaders, innovators, and policymakers to come together and discuss the guiding principles that will shape the future of AI.

We applaud the Korean government's efforts to work together with the global community to develop AI safety governance and policies that are globally coherent and foster a safe but collaborative approach to AI. Korea has a unique opportunity to take full advantage for global leadership in AI Safety governance in the upcoming 2024 Seoul AI Safety Summit by imploring and inspiring other nations, particularly in the Asia-Pacific region, to

come together and cooperate for international regulatory coherence, as the world seeks to ensure safer governance of AI. This would help to foster trust, collaboration, and innovation across borders and sectors, and enable Korea to become a hub in the Asia-Pacific region for AI safety governance.

Article 28 of the 'Digital Bill of Rights,' which implores nations and companies to come together and cooperate to create universal digital norms and mechanism, could be leveraged, and specified as guiding principles for AI safety governance of highly advanced AI, i.e., Frontier AI.

For example, Governments, Companies, Civil Societies, and Academia should:

1. Work together in support of one another to develop universally coherent AI safety standards. These standards would provide guidance and benchmarks for ensuring the quality, reliability, and security of AI systems and their outcomes.
2. Increase investments in developing evaluations for highly capable AI and foster partnerships for sharing best practices around how to develop and conduct these evaluations. Evaluations are essential for assessing the performance, impact, and risks of AI systems and ensuring that they align with the intended goals and values. By sharing best practice, countries can learn from each other and improve their evaluation methods and frameworks.
3. Work to develop and implement policies for the identification, assessment, and management of risks related to highly capable AI models.

4. Work to enhance transparency concerning the capabilities and risks of AI models and systems, and the policies and practices to ensure safety.
5. Continue to evaluate and improve internal governance policies to make AI safer and more transparent. This would include compliance monitoring, and clear delegation of related roles and responsibilities for internal governance on AI safety.

In closing...

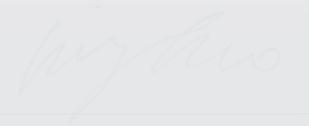
I would like to highlight that Microsoft is honored to be working with the National Information Society Agency (NIA) in developing this paper and hope it serves to demonstrate our commitment to the core principles embedded in Korea's 'Digital Bill of Rights' as well as Microsoft's 'AI Access Principles' which echo and resonate deeply with one another. This partnership and our continuing activities in the global AI domain emphasize Korea's position as a visionary leader in influencing global norms in AI governance.

We invite you to explore this whitepaper—a roadmap toward mutual prosperity through responsible AI governance. This whitepaper provides a comprehensive exploration of Korea's strides in AI governance, reflecting our ongoing efforts to promote responsible, inclusive, and sustainable technological innovation on a global scale.

Together, let us navigate the digital frontier, ensuring AI serves humanity's best interests.



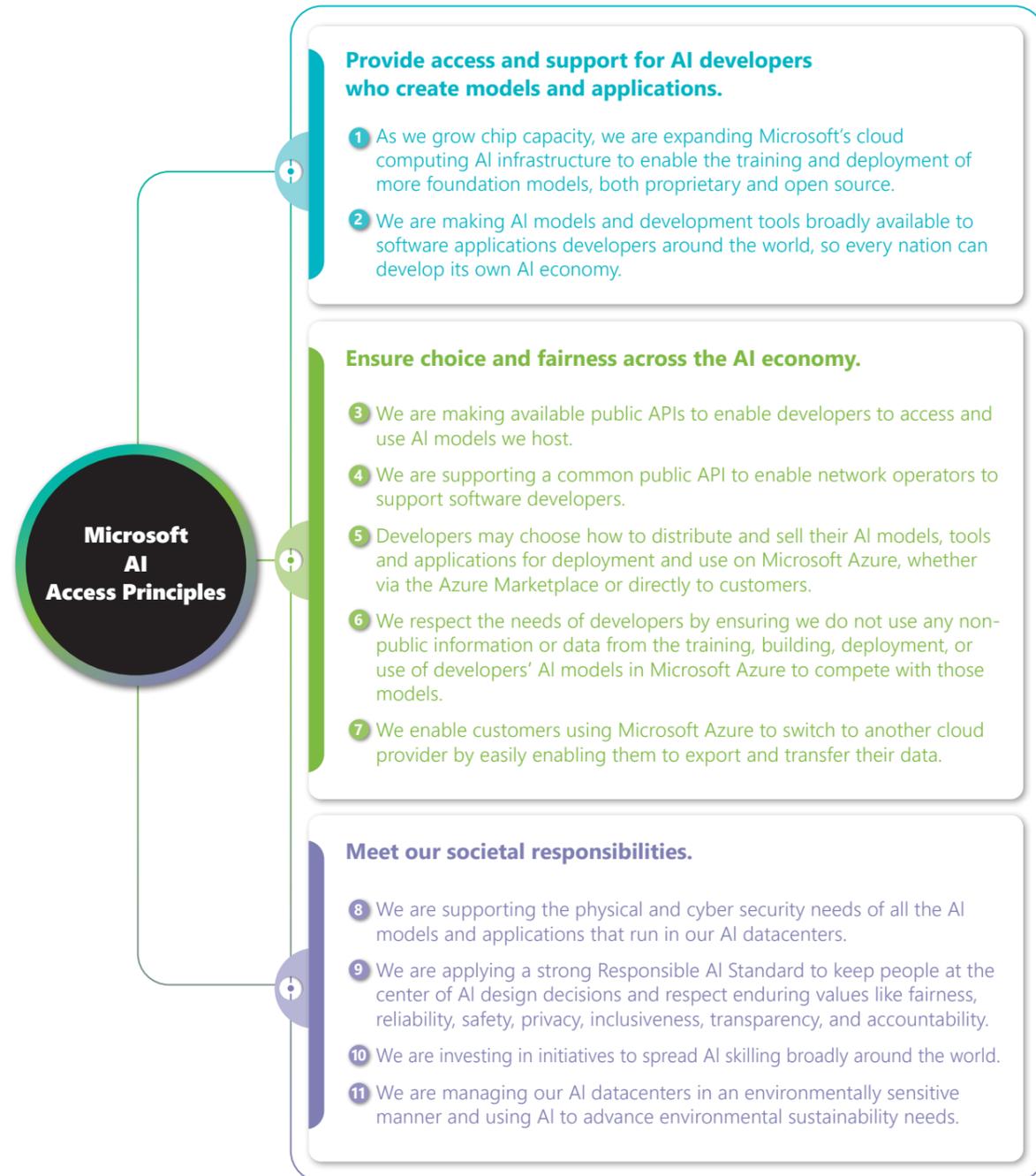
Willy Cho,
Area Vice President,
Microsoft Korea



Willy Cho,
Area Vice President,
Microsoft Korea

Part 1

The development and utilization of Artificial Intelligence (AI) through responsible AI governance is a crucial element of a digital society of mutual prosperity



Microsoft's new AI Access Principles represent a significant step to promoting responsible AI innovation and enhancing accessibility while addressing critical concerns around safety, security, and privacy. These principles are organized into three pillars: providing access and support for AI development, ensuring choice and fairness, and meeting societal responsibilities.

Microsoft recognizes the importance of ensuring widespread access to AI technologies while providing necessary support for developers and users. Under this pillar, the AI Access Principles commit to expanding cloud computing and AI infrastructure, facilitating the development and deployment of AI models through platforms, tools, and services like Azure, and partnering with and supporting other developers, both large and small. By democratizing access to AI tools and resources, Microsoft aims to empower diverse communities to contribute to AI-driven economic growth and collaboration around the world.

Central to Microsoft's AI Access Principles is the commitment to promoting choice and fairness in AI systems. Microsoft pledges to offer public APIs for accessing AI models hosted on Azure, support common public APIs for network operators, and provide developers with options for distributing and selling their AI models. By promoting transparency, accountability, and fairness, Microsoft aims to foster a competitive and equitable AI ecosystem.

Finally, Microsoft recognizes its societal responsibilities as a leader in AI innovation and technology and commits to ethical AI development and deployment by safeguarding AI applications' physical and cybersecurity, adhering to a Responsible AI Standard, investing in AI skilling programs, and advancing environmental sustainability. These efforts underscore Microsoft's commitment to contributing positively to society by aligning AI initiatives with principles of social responsibility, sustainability, and ethical governance.



In the recent evolution of AI governance, various principles have marked a pivotal moment, each further advancing the collective drive to ensure accountability and responsibility of the entire AI ecosystem. As a proactive leader and innovator in AI, Korea is not an exception to this trend and has been establishing principles and a comprehensive AI governance framework that support the growth of AI industry while addressing societal concerns. Korea's such AI policy progress shares a lot in common with Microsoft's principles of fostering AI innovation while ensuring ethical and responsible AI development and deployment. This chapter illustrates Korea's efforts to balance innovation with ethical considerations in the era of AI, with which Microsoft's emphasis on accessibility, fairness, and societal responsibility resonates.

The AI industry has emerged as a core industry for Korea to achieve the goal of coexistence and mutual prosperity.

As early as 2019, Korea recognized the tremendous potential of AI, prompting the government to publish its first National Strategy for AI to revitalize Korean economy, reshape Korean society, and position Korea as a contender among global AI leaders.ⁱ Following the introduction of the Digital Strategy of Korea in September 2022, the Yoon administration has continuously devised a diverse range of strategies to propel AI technology and industry development. These strategies recognize AI's potential across strategic, diplomatic, scientific, and economic domains.ⁱⁱ For instance, the government identified AI as one of the 12 National Core Technologies, considering it essential infrastructure, the growth of which could have similar consequences as the internet or smartphones did for the advancement of society's technology, economy, and diplomatic security. AI therefore is expected to play a significant role in the government's New Growth 4.0 Strategy to drive economic growth.



The Competitiveness Enhancement Strategy for Hyperscale AI introduced in April 2023 also put forth plans to galvanize extensive AI integration into everyday life. This ambitious goal was predicated on Korea's globally recognized digital competitiveness. According to the 2022 IMD World Digital Competitiveness Ranking, conducted by the International Institute for Management Development (IMD), Korea placed 8th out of 63 countries and stood second among Asia-Pacific countries, trailing only Singapore.ⁱⁱⁱ In August 2023, the Ministry of Science and ICT (MSIT) announced its plans to invest KRW 733.1 billion (USD 564 million) in AI-related R&D for Fiscal Year 2024, strengthening Korea's technological sovereignty in this domain.^{iv} Research suggests that the AI industry in Korea is poised for an annual growth rate of 14.9% over the next five years, reaching KRW 4.46 trillion (USD 3.4 billion) in value by 2027.^v Furthermore, a recent report jointly published by Microsoft and the Korea Chamber of Commerce and Industry (KCCI) estimated that generative AI would be able to potentially unlock KRW 620 trillion (USD 476.3 billion) of productive capacity, equivalent to a quarter of the national GDP in 2022, by innovating the workforce across various industries from manufacturing to finances and education.^{vi} As such, the Korean government is committed to sustaining this growth momentum as it attempts to address potential risks effectively, in ways that ensure "[AI's] coexistence with humans is possible."^{vii}

Considering the vitality of the AI industry to Korea's national prosperity, AI governance is moving higher up on the agenda of various government bodies and agencies. As such, Korea has demonstrated proactive yet deliberate steps to address the risks attached to AI effectively, without stifling the momentum for growth. This viewpoint was recently echoed by Minister Lee Jong-Ho of MSIT at the UK AI Safety Summit in November 2023.^{viii} He iterated that the role of policymakers was to secure an appropriate level of AI trustworthiness and safety without hindering AI development and usage. With the MSIT as the central coordinating ministry, relevant ministries and agencies are actively delineating their responsibilities in AI governance through voluntary guidelines and strategies aimed at promoting responsible technology use.

Central to this approach is pursuing the delicate balance of responsible regulation, which leverages earnest efforts to communicate with experts in academia and industry and reflect their voices. Korea has refrained from imposing new regulations or mandates, opting instead to integrate the components of AI governance into existing frameworks and systems. As one of the most AI-ready countries in the Asia-Pacific,^{viii} Korea stands as a prime example of a nation committed to fostering innovation while safeguarding its citizens and the AI industry itself through a thoughtful and adaptable governance framework. This approach highlights Korea's commitment to fully leveraging AI's potential, while prudently managing its implications, contributing to the nation's technological advancement and economic growth.



The Korean Digital Bill of Rights serves as an appropriate starting point to discuss establishing secure and trustworthy AI governance.

At the same time, the launch of the Digital Bill of Rights by MSIT in September 2023 has provided Korea with a meaningful reference for shaping their approaches. These measures seek to establish initial boundaries in the absence of a statutory foundation for AI governance and serve as pointers while the National Assembly of Korea deliberates on a comprehensive legal framework for the nation's AI policy.

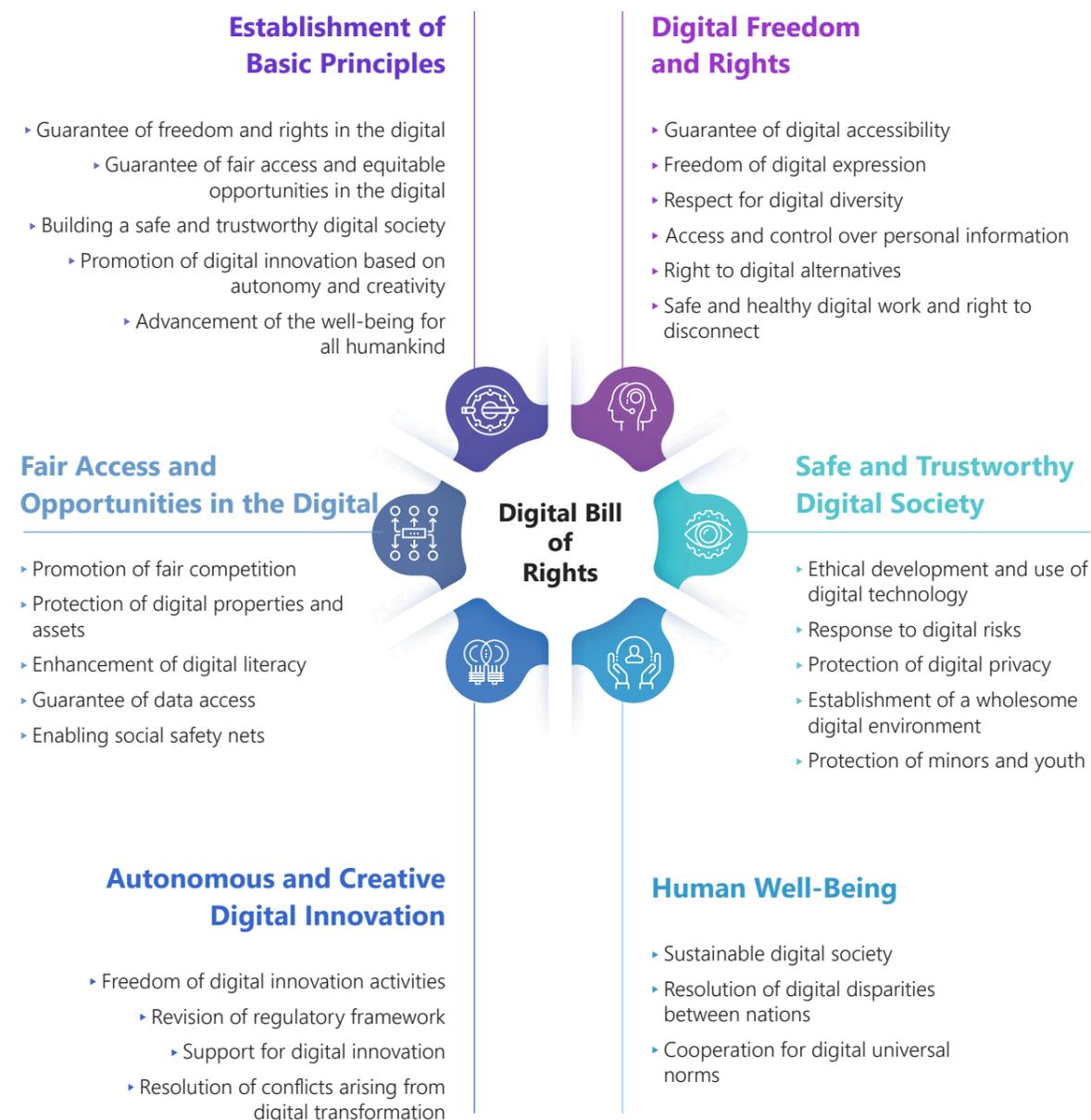
At the Digital Vision Forum held at New York University in September 2023, President Yoon outlined the five basic principles of freedom, fairness, safety, innovation, and solidarity under the Digital Bill of Rights.^x These principles, inherently universal and internationally acceptable as norms, have since been promoted at various international forums, including the UN Global Digital Compact Asia-Pacific regional meeting in October and the OECD Digital Rights Workshop in November 2023.^{xi} In November 2023, at the AI Safety Summit held in the UK, President Yoon also invited global leaders to accept and recognize the five principles as the core values of the AI and digital driven world.^{xii} These global advocacy efforts underscore Korea's commitment to assuming a leadership role in setting norms as a prominent global player in AI.

On September 25, 2023, MSIT formally introduced the Charter for the Values and Principles for a Digital Society of Mutual Prosperity: Digital Bill of Rights.^{xiii} This document serves as a guiding framework for regulators and private companies in navigating the terrain of the digital age, where AI and technology are rapidly reshaping society. Despite its nonbinding nature, the charter's role is akin to a constitution for digital rights, setting the foundational principles for policymakers and companies to build a digital world where innovation is coupled with fairness and trustworthiness. As such, the government will prioritize incorporating these values when harmonizing existing regulations and formulating new ones, such as the Act on Promotion of AI Industry and Framework for Establishing Trustworthy AI (AI Act) and the Digital Inclusion Act¹.

Unlike similar initiatives worldwide like the Blueprint for an AI Bill of Rights by the US, Korea's Digital Bill of Rights is more comprehensive. It addresses a wide range of digital issues beyond AI, encompassing topics such as digital literacy, disparities, and advocating for international solidarity and cooperation. Through this charter, Korea is spearheading a global move to champion a principle-based school of thought to managing and governing the socioeconomic consequences of not just AI but also any future technological advancements. The government aims to promote this approach as a normative model for other nations to adopt, which elucidates the charter's broad, sector-agnostic, and principle-based

nature. The Korean government has consistently emphasized the need for global consensus on digital ethics and AI governance, advocating for a new paradigm that could foster mutual prosperity in our digital-intensive society. The Digital Bill of Rights represents Korea's proposal for such a paradigm.

The Digital Bill of Rights is organized into six chapters, codifying a total of 28 articles that stipulate high-level principles under five essential criteria. The following diagram provides an overview of the blueprint for a digital society of mutual prosperity set out by the Digital Bill of Rights: Domestic experts commend Korea's balanced approach to AI governance, which effectively tackles the immediate and interconnected risks posed by digital technologies through principle-based guardrails. These guardrails



¹ Digital inclusion refers to the improvement of accessibility that allows everyone to use IT devices and services and acquire information without barriers. As of December 2023, two proposals for the Digital Inclusion Act, each submitted in January 2021 and November 2022, respectively, await subcommittee review in the National Assembly.

enable timely and adaptable responses from the regulatory and private sectors. Such careful approach, so as not to jeopardize the momentum of innovation, stems from Korea's unique socioeconomic structure and its position as a small yet strong global tech leader with a burgeoning AI industry to promote and valuable expertise to share.

Being one of the world's largest exporting countries with IT and manufacturing industries of a high caliber, it is in Korea's utmost interest to align with global standards and regulatory movements. Moreover, active participation in international discussions is crucial to ensure that domestic companies and industries are not disadvantaged by external regulations that arise. By operationalizing the high-level principles enshrined in the Digital Bill of Rights and continuing to exchange those values and practices on the global stage, Korea has the opportunity to further its normative relevance and influence in AI governance. This, in turn, would bolster the global competitiveness of the Korean economy in the digital society and help prevent potential trade disputes.

"[In terms of global AI governance discussions,] the priority for Korea is to have a seat at the table, and if possible, at the head table. This will be achieved when we are able to valuably contribute to establishing the basic framework that undergirds governance systems constructed by individual countries. Here the goal isn't to fashion global norms that comply with or further Korea's own prerogatives, but to ensure that whatever framework emerges, it is both universal and at the same time compatible with our societal values, culture, and economic objectives, thereby allowing Korean industry and actors to continue to thrive in the global economy."

- Prof. Lim Yong, Director, Seoul National University AI Policy Initiative

Therefore, the reasonable next steps for Korea would involve examining the implications of these core principles in the context of AI governance, specifying the issues and risks to which they should be applied, and deliberating on the underlying societal ramifications that the governance framework aims to manage by mitigating the identified risks. In the process of doing so, specific measures that implement these principles would need to be developed in such a way that safeguards people without depriving them of the opportunity to take advantage of AI advancement.

With this knowledge, the following chapter pursues how each of the five principles for a mutually prosperous digital society has so far been integrated into various efforts and advancements in the discussion of AI governance at three levels: within Korea, internationally and in the private sector.

Part 2

Korea, the international society, and the tech industry are on an organic journey towards a digital society of mutual prosperity



Korea’s principle-based approach to AI governance for the goal of a digital society of mutual prosperity is in line with the response of the international community and the tech industry, which argues for the development and use of AI under the three themes of responsible AI, inclusive AI, and sustainable AI.

Korea’s approach to AI governance aligns with global consensus and reflects a commitment to fostering a mutually prosperous digital society. The nation has adopted a principle-based framework for the development and deployment of AI technologies. This approach emphasizes the importance of responsible considerations, accountability, and societal well-being in the advancement of AI applications.

In parallel with the international community and the tech industry, Korea’s AI governance strategy is built upon three overarching themes: responsible AI, inclusive AI, and sustainable AI. The nation’s commitment reflects a holistic approach that prioritizes the well-being of society while embracing the transformative potential of AI.

Considering the broad, value-based nature of the three themes for a mutually prosperous digital society, the implications of each principle within the realm of AI policy and governance may initially appear ambiguous. This chapter highlights how these themes are manifested in diverse responses and initiatives and driven by the community and industry both within Korea and globally. The mapping process used in this chapter between the principles of the Digital Bill of Rights, their role in AI governance, and the value to society, draws insights from the following key international developments selected for their current relevance and potential impact on shaping global norms for AI governance:

- 01 US White House Executive Order on the Safe, Secure, and Trustworthy AI
- 02 US National Institute of Standards and Technology (NIST) AI Risk Management Framework
- 03 UK AI Safety Summit and the Bletchley Declaration
- 04 G7 Guiding Principles for Developing AI
- 05 OECD AI Principles
- 06 EU AI Act

Through a comprehensive comparative analysis, the significance of the Digital Bill of Rights in the context of AI governance is elucidated, particularly as it aligns with the themes of responsible AI, inclusive AI, and sustainable AI. This analysis not only underscores the alignment of the Digital Bill of Rights with the evolving landscape of AI governance principles but also provides valuable insights and potential avenues for improvement. This information will be particularly beneficial for Korean policymakers who are actively involved in the formulation and refinement of the AI governance framework. The exploration of this comparative perspective serves as a guiding resource, empowering policymakers to make informed decisions and enhance the effectiveness of AI governance in the Korean context.

Korea’s approach to responsible AI

Responsible AI entails the ethical and conscientious use of AI technologies, emphasizing fairness, transparency, and accountability. This principle ensures that AI systems are developed and deployed with due consideration for their societal impact, minimizing biases and promoting equitable outcomes.

The concept of responsible AI within Korea’s Digital Bill of Rights highlights the necessity of adopting and implementing AI technologies in a manner that prioritizes responsible practices. Principles 3 and 4 calls for the creation of a safe and trustworthy digital society by incorporating fairness, transparency, and accountability into the development process, thereby mitigating biases, and ensuring that AI applications benefit society while promoting digital innovation. The commitment to responsible AI reinforces the idea that technological advancements should go hand in hand with responsible considerations to build trust and promote positive societal impact.

Digital technologies and services shall not pose a threat to the safety of individuals and society. Appropriate means and procedures must be in place to prepare for and respond to digital risks. This principle is further expounded by five provisions below:^{xiv}

**Principle 3
Creation of a safe and trustworthy digital society**

<p>Ethical development and deployment of digital technology</p> 	<p>The development and deployment of digital technologies must be carried out responsibly in an ethical manner to ensure safety and trust.</p>
<p>Response to digital risks</p> 	<p>Digital risk should be prevented and managed through appropriate measures and procedures, and information about the risk should be accessible and transparently disclosed to the public.</p>
<p>Protection of digital privacy</p> 	<p>An individual’s privacy must be protected from illegal identification and tracking, including digital surveillance, location tracking, etc.</p>
<p>Establishment of a wholesome digital environment</p> 	<p>A healthy digital environment should prevent manipulation based on fake news and the production and distribution of illegal and harmful information, and effective measures and procedures should be in place to protect victims from digital crimes that occur.</p>
<p>Protection of minors and youth</p> 	<p>Children should have the freedom to participate in an age-appropriate digital space while being afforded additional protection from potential crimes that may arise from digital technologies.</p>

Within the context of AI governance, these provisions, especially the first three on **ethical development and deployment, Digital risks, and Privacy**, have a direct linkage to discussions about AI policies. The table below illustrates how they may be anatomized in the context of AI governance:

Provision	Contextualization in AI governance	Underlying societal concerns
Ethical development and deployment of digital technology	<ul style="list-style-type: none"> • Cooperation and shared accountability across developers, deployers, and operators of AI systems throughout AI lifecycle • Development of sector-specific and stage-specific guidelines • A mechanism for transparent information disclosure • An impact assessment scheme • A licensing scheme for highly advanced models 	<ul style="list-style-type: none"> • Bias and discrimination, social unrest, workforce displacement, public trust
Response to digital risks	<ul style="list-style-type: none"> • Risk control and management measures • A case-by-case evaluation scheme • Risk-based ex-post penalties • Equipping people with relevant skillsets and capabilities to ensure reliable operation • Public-private platform for knowledge sharing and best practices • A national registry of high-risk AI systems 	<ul style="list-style-type: none"> • Exploitation of AI systems for malicious purpose such as hacking, phishing, and malware, digital inclusion (protection of digitally illiterate citizens), national security, hybrid or cyber war
Protection of digital privacy	<ul style="list-style-type: none"> • Robust data security measures and a clear reporting structure in case of a breach • Regular audits and assessments to identify privacy risks and take corrective actions when needed • Compliance with relevant data protection laws and ethical standards 	<ul style="list-style-type: none"> • Individual autonomy, legal and ethical obligations, misuse of technology for malicious purpose such as stalking



Korea

In Korea, responsible AI is one of the core topics around which domestic discussions have taken place. In December 2020, the MSIT launched the first **National AI Ethical Standards** and underscored the values of privacy protection, safety, and transparency, among others, as key requirements for AI. Based on these standards, the MSIT is in the process of supporting the industry in interpreting these standards in daily practices and processes. In tandem, it is also identifying regulatory or structural challenges that hamper such application. In February 2022, the MSIT launched the **AI Ethics Policy Forum**, composed of experts from academia, industry, and civil society. In collaboration with the Korea Information Society Development Institute (KISDI), a government-affiliated research institute in ICT policy, the Forum produced a **self-auditing checklist** and **development guidelines for companies**, as well as **teaching materials** and **instructional guides for AI ethics education**.

Furthermore, the MSIT is developing a **framework for private sector-led AI reliability and ethics regulation**, along with a **management system for assessing AI impact**. Following up from the first phase AI Ethics Policy Forum, the second phase launched in April 2023 aims to create **sector-specific self-checklists** and **guidelines designed for AI system developers and operators** in areas where reliability and ethics are emphasized, such as recruitment, public safety, and services using generative AI by December 2023. The **impact assessment system** will introduce voluntary compliance criteria for risk management, responsible compliance, and the impact of AI products and services.^{xv}

“For responsible AI to spread throughout society and take root, there needs to be a balanced triad of inclusive technology development, developer/deployer ethics, and user ethics. This calls for all stakeholders to collectively contemplate and prioritize human-centered approaches to realize a safe society from the risks of digital technology.”

- Choi Moon-sil, Director, National Information Society Agency (NIA) Department of Digital Inclusion

Additionally, in collaboration with the Telecommunications Technology Association (TTA), the only institute in Korea dedicated to establishing, revising, and disseminating ICT-related standards, the MSIT has formulated **guidelines for the development of AI systems**. These guidelines encompass a testing and certification framework designed for AI system developers and operators. Although voluntary in nature, the MSIT and TTA aspire to utilize this framework to instill a standardized trustworthiness-first approach within Korea's AI industry. This will be achieved through the provision of consulting and training in the application of the framework.

For MSIT-led research and development projects, participating companies and organizations may be required to pursue certification as part of the requirements. The **safety and trustworthiness testing and certification system**, initiated in October 2023, is presently undergoing a trial period until December 2023.^{xvi}

"Similar to a knife, and like everything else we use in our daily lives, AI also requires continuous investment and dedication in technological development. However, it is crucial to provide guidelines to prevent the misuse or abuse of the developed technology. ... TTA has created a development guide for the implementation of four technically achievable principles out of the 10 National AI Ethical Standards - diversity, stability, responsibility, and transparency - in AI services and products. By creating such development guides and providing training and consulting, we are fostering a system for companies to autonomously consider fundamental requirements for reliability from the product and service planning stages, thereby guiding companies to [conform to requirements for the development of safe and trustworthy AI systems]. ... It is expected that a standardized reliability verification system will contribute to the growth of the AI industry."

- Lee Kang-hae, Team Leader, TTA AI Convergence Planning Team

The National Assembly of Korea has also been keen to integrate considerations of AI responsibility, fairness, trust, and transparency into existing regulatory frameworks through legislative amendments, the ongoing efforts can be found in Appendix A. The pending **AI Act** takes a risk-based approach by scoping out high-risk AI systems that should be met with more stringent disclosure requirements.^{xvii} Also noteworthy is the ex-post regulatory approach that allows AI technologies to enter the market with relevant changes in regulations to follow, as necessary, which reiterates the Korean government's keenness to facilitate industry growth.

Overseeing data privacy issues in AI governance, the Personal Information Protection Commission of Korea (PIPC) has committed to a transition from a rule-based to a principle-based regulatory framework with differential compliance measures based on the valuation of sensitivity and risks.^{xviii} These efforts are enabled through public-private partnership, such as the **AI Privacy Public-Private Policy Council**, where robust and practical standards concerning privacy and data protection in AI systems are to be developed.

In summary, Korea's current endeavors in AI governance, a building block for a safe and trustworthy digital society, reflect a risk-based approach that embraces self-regulatory initiatives led by the private sector.

Global developments

This direction taken by Korea in AI governance is mostly in line with that of the international society. Security, safety, and risk management have become recurring themes in the AI governance models of other countries and international principles. The OECD AI Principles formulated in 2019 which Korea built its National AI Ethical Standards upon, suggests robustness, security, and safety as one of the value-based principles for the development and deployment of trustworthy AI.^{xix}

Another notable derivative of the OECD AI Principles, the recent G7 International Guiding Principles on AI emphasizes a risk-based approach. It encourages implementing measures to identify, evaluate, and mitigate risks in the development of advanced AI systems, as well as after the deployment on the market.^{xx} Specifically, Principle 5 explicitly calls for AI governance and risk management policies grounded in a risk-based approach, while Principle 6 mandates the establishment of security controls across the AI lifecycle.

A distinctive departure from the risk-based approach found in Korea involves an additional layer of a lifecycle-based approach. This can be perceived as an effort to account for the diverse uses of and interactions with AI systems by stakeholders, acknowledging corresponding differences in risks and responsibilities that each stakeholder needs to prioritize.

The White House Executive Order on the Safe, Secure, and Trustworthy AI (the White House Executive Order) also entails standards for AI safety and security to guard against the potential risks of AI systems.^{xxi} These requirements include developing standards, tools, and tests to ensure the safety of AI; investing in an advanced cybersecurity program to spot vulnerabilities in critical software; classifying "the most powerful" or high-risk AI systems and mandating developers of such AI models to share safety test results and critical information with the government; and formulating guidance on content authentication to identify AI-generated content and protect citizens from AI-enabled fraud. The Executive Order also goes at length to specify how the government will support privacy-preserving research and technologies so that privacy risks posed by AI might be appropriately mitigated.



In addition to the White House Executive Order, the US has also produced, with the National Institute of Standards and Technology (NIST) leading the effort, the AI Risk Management Framework (AI RMF) in January 2023 which emphasizes AI risk controls.^{xxii} Developed in close collaboration with private and public sectors, the AI RMF provides an overview of three types of AI Harms and outlines seven characteristics of trustworthy AI systems. Similar to the lifecycle-based approach manifest in the G7 AI Principles, the AI RMF recognizes the AI lifecycle as a crucial point of consideration for effective management of AI risks.^{xxiii}

In November 2023, over 150 representatives from governments, industry, and civil society gathered at the UK AI Safety Summit, which culminated in the Bletchley Declaration signed by 28 countries.^{xxiv} The Declaration reached an initial agreement on frontier AI and expressed the need for national and international cooperation on AI risk identification and mitigation. In doing so, the Declaration called for countries to adopt risk-based policies and build legal and regulatory frameworks based on national circumstances. In furtherance of this goal, countries should aim to build “a scientific and evidence-based understanding of these risks,” which may involve utilizing impact assessments and testing tools.^{xxv}

As such, ensuring the security, safety, and trust of AI technologies stand as paramount objectives of AI governance. Numerous global initiatives indicate a preference for a risk-based approach that accounts for the diversity in the AI value chain. This approach enables the development of more tailored fit compliance measures based on one’s role, whether as a developer of a large-scale AI model, a deployer of a service using the said model, or both.

Private Sector

Some leading AI companies, such as Microsoft, have been extensively collaborating with different stakeholders to operationalize AI safety and security, demonstrating trustworthiness and responsibility. Regarding the principle of facilitating a safe and trustworthy digital society, Microsoft acknowledges both the potential of AI to help make society safer, and the potential risks if the technology is used adversely, or without appropriate care.

In the context of the Russia-Ukraine war, Microsoft proactively tracked and monitored Russia’s cyber threats as early as January 2022, preceding the latter’s formal invasion.^{xxvi} Microsoft’s world-leading AI algorithms and analytics facilitated the identification of patterns, behaviors, and potential vulnerabilities in real-time. This information was then communicated to the Ukrainian government through AI-driven 24/7 threat intelligence platforms that enabled the rapid dissemination of actionable insights and the corresponding deployment of countermeasures. Through the effective utilization of AI and cybersecurity capabilities, Microsoft identified and mitigated digital risks associated with cyber threats, exemplifying how a committed and responsible private actor can contribute to the

international cooperation in creating a safer and more resilient digital society. Drawing from its wealth of knowledge and experience like this, Microsoft has also been producing valuable resources for the benefit of the global security community.²

Considering Korea is still at war with North Korea, the heightened risk of AI-intensified cyberterrorisms poses a real threat to the Korean government that must be averted. To ensure Korea’s national security, maintaining cybersecurity capabilities of the highest caliber in the face of complex and dynamic cyber threats is of utmost importance, which can be supported by the exponential power of AI.

Simultaneously, the power of AI must be harnessed and managed responsibly, a principle that Microsoft keenly abides by. Prior to the issuance of the White House Executive Order, seven leading AI companies, including Microsoft, Google, Amazon, Meta, and OpenAI, committed to a set of Voluntary Commitments with the US government. The objective was to demonstrate the principles of safe, secure and trustworthy development of AI technology.^{xxvii} These commitments encompass internal and external security testing by experts prior to public use, cross-sector sharing of best practices, investment in cybersecurity and insider threat safeguards, as well as robust reporting and discovering by third parties. Importantly, these commitments align with AI governance measures endorsed in global principles. The securing of this first wave for voluntary commitment from AI companies had attracted additional companies to sign on.^{xxviii} The expressed support from the private sector had laid a sound foundation for the US Government to push forward with the Executive Order on AI.

Microsoft has consistently led by example over time by embracing and embodying new global standards and compliance requirements into its products and practices ahead of the curve, whether it be the EU’s General Data Protection Regulation (GDPR) or the Responsible AI (RAI) movement.^{xxix} In the case of the GDPR, Microsoft was the first hyperscale cloud provider to prepare GDPR terms and conditions at the enterprise level, while it had also initiated the first step to embed responsible AI in its research, governance, and engineering. In the domain of AI governance as well, Microsoft continues to showcase accountability and responsibility. Particularly impactful about Microsoft’s thought leadership in AI governance is that, as one of the largest providers of digital infrastructure, Microsoft’s efforts translate into the safety, security, and trust of AI development and deployment for its clients across sectors and industries. This, in turn, has strengthened the entire AI ecosystem, making it that much more resilient.

² Microsoft’s flagship products, showcasing recent trends in cybersecurity and providing actionable insights for all stakeholders to contribute to a safer digital society, include the Microsoft Digital Defense Report (now in its fourth annual edition) and Digital Front Lines, a special report developed in partnership with FP Analytics with a focus on hybrid warfare. These can be accessed here: <https://www.microsoft.com/en-us/security/security-insider/microsoft-digital-defense-report-2023>; <https://digitalfrontlines.io/>

Principle 4
Promotion of autonomous and creative digital innovation

Principle 4 stipulates that a digital society is one in which the ongoing development of digital technology and innovation is encouraged. This culture of innovation should be built upon a foundation that respects individuals' autonomy and creativity. In the Digital Bill of Rights, four provisions are suggested to foster such culture:^{xxx}

<p>Freedom of digital innovation activities</p> 	<p>Everyone is entitled to the freedom to engage in economic, social, and cultural activities that promote digital innovation in diverse fronts.</p>
<p>Revision of regulatory framework</p> 	<p>A reasonable regulatory system that respects civil liberty for the promotion of digital innovation and the regulation should be reasonably reformed in consideration of social acceptability, industrial maturity, and the speed of technological development.</p>
<p>Support for digital innovation</p> 	<p>Professional workforce training, research and development investment, and start-up revitalization, as well as infrastructural development and systems building should take place in close cooperation between the private and public sectors for the continuous creation of digital innovation.</p>
<p>Resolution of conflicts arising from digital transformation</p> 	<p>An institutional base for multistakeholder discussions and deliberation should be made to manage and prevent social conflicts that may arise in the process of digital transformation and reach a societal consensus.</p>



One area where the Korean government's efforts have been prominent and particularly relevant to the discourse of AI governance would be **Revision of regulatory framework** to abolish outdated regulations that no longer fit with the digital environment. This provision is already being put into practice through various measures outlined below to reform and update the existing regulatory framework to encompass the AI industry, thereby reinforcing the basis for a globally competitive digital society in a responsible manner.

Provision	Contextualization in AI governance	Underlying societal concerns
<p>Regulatory reform and revision</p>	<ul style="list-style-type: none"> • Amendments to existing regulatory frameworks such as data governance, financial systems, etc • Expansion of the post-regulatory (blacklist) approach to permit innovation • Expansion of the risk-based approach • Detailed guidelines to complement ex-post regulations and provide legal clarity and certainty • Public consultations and forums for discussion • Regulatory sandboxes to test and deploy innovative AI applications • Cross-border reference building and sharing • Education and training for regulators • Incentives to promote self-regulations by the private sector 	<ul style="list-style-type: none"> • Lagging behind in a digital society, exacerbation of digital divides and inequities, unauthorized use and exploitation of personal data, and unmet ethical considerations



Korea

From the outset, Korea has refrained from imposing new regulations or mandates, opting instead to integrate the components of AI governance into existing frameworks and systems. In May 2020, the MSIT, together with the National Information Society Agency (NIA), first launched the **AI Legislation and Regulation Committee**, tasked with collecting opinions on social changes likely to be caused by AI and developing an agenda to reconcile existing laws, regulations, and interpretations by regulators and judicial authorities with those anticipated changes.^{xxxi} This whole-of-society committee consisted of members from the public sector, such as the MSIT, the Ministry of Justice, Ministry of Culture, Sports and Tourism, and the Ministry of Interior and Safety, along with private members from industry, business, law, and academia. In December 2020, the committee unveiled a **roadmap outlining 30 jointly identified tasks** that should be prioritized to strengthen the foundation for the development and utilization of the AI industry and prevent adverse effects.^{xxxii}

A noteworthy aspect of the roadmap's development was the careful consideration of Korea's domestic legal systems vis-à-vis overseas legislative trends for global coherence. The roadmap emphasized that regulatory reform and revision should foster industry autonomy through

self-regulation and be grounded in social consensus to carry weight and achieve the desired impact.

This roadmap supported discussions within the Korean government about amending the **Personal Information Protection Act (PIPA)**. Following its enforcement in September 2023, the amended PIPA will now grant data subjects the right to refuse decisions made by automated systems for processing personal information if these decisions significantly impact their rights or obligations. It also provides data subjects with the right to request explanations for these decisions.^{xxxiii} Additionally, it prompted discussions on amending the Copyright Act to enable large data use in analysis and AI learning, although this amendment is still pending.

In August 2023, the **4th AI Legislation and Regulation Committee** was launched, this time to update the initial AI legislative roadmap and enhance regulatory frameworks for hyperscale AI.^{xxxiv} Structured into four specialized subcommittees addressing regulatory enhancements, trust-building measures, sector-specific considerations, and AI-related copyright matters, the committee will clarify the definitions and scope of key terms in AI governance, such as high-risk AI. It will also evaluate Korea's alignment with global AI standards and identify new areas of reform. The updated roadmap for regulatory reform is expected this year.

While the National Assembly of Korea deliberates on the comprehensive legal framework for the nation's AI policy through the draft AI Act, it has concurrently been incorporating AI governance into Korea's existing regulatory frameworks through legislative amendments.^{xxxv} Introduced bills for amendment range from the **Fair Hiring Process Act** and **Content Industry Promotion Act** to the **Information and Communications Network Act**, manifesting the Korean government's awareness of and inclination to manage AI and its societal impact in the most efficient manner possible.^{xxxvi}

In addition to revising existing regulations for industry autonomy and innovation, the Korean government employs the **ICT Regulatory Sandbox** to waive regulations to a limited extent.^{xxxvii} Identifying gaps between new technologies and the current regulatory structure, the Sandbox tests and verifies the safety and impact of the technology, allowing the government to monitor potential risks and gather insights for regulatory improvement. Currently, two AI system deployers are demonstrating their services in the Sandbox, undergoing safety testing.^{xxxviii}

Interviews with experts highlight global consistency and coherence as a crucial aim for the Korean government, which exercise influence over the development of national strategies and policies, such as the Digital Bill of Rights, the Ethical Standards and the PIPC's policy directions for regulating use of personal information in AI. This emphasis on global coherence corresponds to the interconnectedness and immediacy of digital transformation. To transition to a mutually prosperous digital society,

harmonizing the national AI governance framework with that of the global has become essential. The harmonization also helps in ensuring the domestic AI industry develops and becomes secure in tandem with global trends. In line with this reasoning, relevant government institutes like NIA, the TTA, and KISDI are all committed to review and ensure alignment of the Korean approach with international standards.

Overall, Korea's approach to AI governance reflects a commitment to innovation and industry autonomy, responsiveness to the evolving landscape of AI, and awareness of the overarching transformative impact that AI will have on Korean society. The collaborative efforts of the AI Legislation and Regulation Committee, coupled with initiatives like the ICT Regulatory Sandbox, showcase the country's dedication to refining its AI governance framework to align with global trends, and strengthen the industry's competitive edge in the global market. These strategic and adaptive measures position the nation at the forefront of fostering a conducive environment for technological innovation and responsible AI development and deployment.

Global developments

Other countries and global partnerships are also aware of the need for responsible regulatory updates to facilitate the responsible development and deployment of AI, as demonstrated by the EU AI Act, the US White House Executive Order, and other guiding principles.

In an effort to establish a framework that ensures user trust in AI, the EU AI Act advocates for the use of regulatory sandboxes.^{xxxix} Within these sandboxes, innovation would be fostered in a controlled environment with oversight, and additional measures would be implemented to ease the regulatory burden for SMEs and startups. This "innovation-friendly" approach has also been endorsed by the French, Italian, and German governments, emphasizing their collaborative approach to AI governance. They remain dedicated to "reducing unnecessary administrative burdens on companies that could impede Europe's ability to innovate," which may involve significant reforms, simplifications, or the facilitation of fast-track procedures to authorize AI innovation and investments in the EU.^{xl}



The White House Executive Order also champions innovation and competition, with one strategy being the assurance of a fair, open, and competitive AI ecosystem.^{xli} For example, the US plans to modernize its immigration policies to attract AI talent, while adopting regulatory measures to grant SMEs and startups access to resources and public assistance for more active participation in the industry. In terms of the role of experimentation, the Executive Order seems to indicate using the adoption of AI by government agencies as a sandbox to monitor and test AI governance for future actions.^{xliii}

A crucial aspect of how the US government approaches regulatory streamlining efforts is the emphasis on inclusion and consultation with industry and relevant stakeholders to accurately establish and modify regulatory guardrails. Particularly when addressing the high risks of AI and its potential implications, developing responsible policy and regulatory approaches, and advancing innovation, the Executive Order mandates the solicitation of input from the private sector.^{xliii} The advisory committee to be established for AI safety and security will include AI experts from the industry and will be expected to engage directly with the Secretary of Homeland Security to ensure the resilience of AI usage in critical infrastructure. Thus, the US' regulatory revision process provides ample space for private sector participation, recognizing that voices from the field are essential for building applicable and appropriate guardrails that fit with the context of national society.

Apart from the above, the OECD AI Principles also urge governments to establish an agile regulatory environment that fosters responsible innovation through controlled experimentation, and the continuous review

and reform of their policy frameworks under Principle 2.3.^{xliv} All of these aspects, as outlined earlier, are evident in Korea's approach to facilitate autonomous innovation in the private sector.

The Bletchley Declaration from the UK AI Safety Summit also acknowledges a pro-innovation approach. It emphasizes the need for governance mechanisms that can effectively maximize the potential of AI in a responsible manner.

In the realm of fostering innovation, the imperative of global alignment becomes evident, as exemplified by the recent developments surrounding the EU AI Act passed in June 2023. Amendments to this legislation have been instituted to harmonize with international standards, particularly those defined by the OECD, aiming to enhance the relevance of their governance framework and ensure active participation in discussions on international standards in the future.^{xlv} In this context, Korea may consider reassessing its existing proposals for AI governance, including the draft AI Act, to pinpoint areas for closer alignment with international practices. An aspect that warrants attention is the more precise categorization of stakeholders within the AI industry. Under Korea's current approach, all participants in the AI ecosystem are amalgamated and acknowledged as business operators.^{xlvi} In contrast, the OECD, EU, and the US employ a more specific categorization by distinguishing between AI system developers and AI system deployers.^{xlvii} This approach takes into account, as previously explained, the diverse roles, risks, and corresponding responsibilities applicable to each industry actor based on their position in the AI value chain.

The principle of innovation and autonomy has been widely echoed by global actors in AI governance, where governments recognize that regulatory frameworks should not undermine industry-led innovation. To strike this balance, there is a trend of organizing controlled experimentation and inviting experts from the industry to gain real-life insights. This approach aims to reform policies in ways that they are more impactful and relevant to enhancing responsible AI development and deployment through innovation.

Private sector

Microsoft advocates for an AI governance framework aligned with the technology architecture and actively engages with the public sector to contribute to successful adaptation of such reform.^{xlviii} As a leading tech industry player, Microsoft has expounded the dynamic structure of the AI ecosystem through thought leadership, including the five-point Blueprint for Governing AI, and participation in relevant discussions taking place at the national and international levels. Given AI's dual role as both an innovation itself and a transformative infrastructure for creative applications, an innovation-friendly regulatory revision necessitates identifying tailored responsibilities based on industry players' diverse roles in managing and utilizing AI technology. Microsoft continues to assist



governments in reflecting such industry characteristics to facilitate effective and efficient regulatory revisions, emphasizing the adequacy of existing measures, especially at the applications level, and deliberating the best approach to develop new ones when needed.

At the vanguard of innovation and creativity in the digital era, Microsoft has exercised prudence in every step, ensuring accountability for the outcomes and impacts of its changemaking. In 2017, amid the escalating rate of cybercrime and the disconcerting proliferation of cyberattacks to an unprecedented level, Microsoft advocated for the convening of the Digital Geneva Convention.^{xlix} While safeguarding and defending its customers at the corporate level, Microsoft had recognized the need for collective, concerted efforts on a global scale and urged governments to unite in addressing this emerging risk. By then, Microsoft had already proactively allocated over USD 1 billion annually to security-related research and investment.

Another example pertains to the management of sensitive AI use cases, particularly facial recognition technology. Since 2018, Microsoft has actively advocated for governments to implement guardrails to prevent the misuse of facial recognition that could violate citizens' fundamental human rights.ⁱ Microsoft presented insights into the potential societal and economic consequences of facial recognition technology, offering recommendations for both the public and private sectors to effectively govern problematic AI use cases. This illustrates Microsoft's commitment to provide a constructive perspective on policy as a responsible corporate member of society. Finally, in 2020, its advocacy bore fruit with the passage of Washington state legislation on facial recognition.ⁱⁱ

Drawing on many years of work, research, and input, Microsoft boldly takes aggressive steps to revise, update, and future-proof its products and services against evolving risks and concerns associated with innovation. This vigilant approach elucidates how Microsoft, as mentioned earlier, was the first to institutionalize the concept of Responsible AI and promptly developed the Customer Copyright Commitment (formerly the Copilot Copyright Commitment) in response to the emerging issue of copyright infringement in generative AI in 2023.ⁱⁱⁱ Notably, the Customer Copyright Commitment itself is a revision, building on the foundation of the AI Customers Commitments. These commitments not only inform Microsoft's customers about the techniques at Microsoft that protect and facilitate their creative autonomy but also provide useful references for other companies and governments.

Microsoft's agile revision of its processes, adapting to evolving needs and societal risks as technology advances, is complemented by proactive information sharing with other stakeholders, contributing to the common good. This capability is facilitated by Microsoft's dedication to monitoring risks in real-time and taking swift action based on a holistic, data-driven analysis of the gravity, severity, and complexity of the issues at hand.

"Big tech companies like Microsoft have established internal systems, including their own ethical principles, technological measures, and research resources to guarantee reliability and trustworthiness [of AI systems]. On the other hand, SMEs and startups, while being interested in innovative development and dissemination of technological advancements, may lack adequate preparation to appropriately address the potential risks of technology. In this regard, we create and distribute AI ethics self-checklists to address these issues, but if big tech companies could share their internal policies, managerial efforts, and institutional initiatives for securing AI ethics and reliability, it is believed to contribute to a healthier growth in the overall industry ecosystem. Particularly, by collecting various best practices and disseminating them to SMEs and startups, it is expected that as more startups successfully integrate and grow by plugging into the foundation models of large companies, mutual growth becomes possible. Therefore, it is thought that big tech companies should take on the role of securing AI ethics and trust."

- Dr. Moon Jung Wook, Director, KISDI Department of Intelligent Information Society Policy Research

Korea's approach to inclusive AI

Inclusive AI underscores the importance of addressing diversity and accessibility in AI technologies. Korea recognizes the need to ensure that AI benefits all segments of society, regardless of demographic factors. This includes efforts to mitigate biases in AI algorithms and promote inclusivity in the development process, allowing for diverse perspectives to be considered and avoiding the creation of technologies that inadvertently marginalize certain groups.

Principles 1 and 2 of the Digital Bill of Rights upholds the theme of inclusive AI by underscoring the need to protect digital freedom, human rights, and by guaranteeing fair access and equitable opportunities in digital society.



Principle 1 Protection of digital freedom and rights in the digital

The first principle of the Digital Bill of Rights emphasizes this theme by establishing respect for human dignity and values as the foundation of the digital society in which everyone's rights and freedoms are guaranteed. This principle explicitly outlines six digital rights and freedoms:^{liii}

Assurance of digital accessibility 	Everyone is entitled to a stable network environment, ensuring that various digital services should be accessible and available at all times and everywhere.
Freedom of digital expression 	Everyone is entitled to freely express their will in a digital environment. However, this right should be exercised responsibly to avoid infringing on others' rights, public morality, or social ethics.
Respect for digital diversity 	Everyone must be protected from unreasonable discrimination and prejudice resulting from the use of digital technology and be respected for their social and cultural diversity.
Access and control over personal information 	Everyone must have access to and control over information about themselves in a digital environment, including, but not limited to, the rights to access, correct, delete, and transfer such information.
Right to digital alternatives 	Everyone has the right to request alternative methods to substitute digital methods in the public sphere.
Safe and healthy digital work and right to disconnect 	Everyone is entitled to work safely and healthily in various work environments arising from the development of digital technology and to ensure rest away from digital connectivity.

While the other human rights and freedoms outlined in Principle 1 are relatively straightforward in terms of enforcement in the context of AI, the **Guarantee of digital accessibility** may seem less familiar. The table below illustrates what ensuring accessibility through AI governance may entail and unpacks the underlying social concerns that make digital accessibility a crucial objective in the AI era:

Provision	Contextualization in AI governance	Underlying societal concerns
Guarantee of digital accessibility	<ul style="list-style-type: none"> Promotion of AI-powered assistive technologies with human-centred design in public service digitalization Accessibility guidelines and standards for developers and deployers of AI systems and services A diversity and inclusion advisory group Training and awareness programs on implementing inclusive features in AI development and deployment 	<ul style="list-style-type: none"> Exclusion of a certain population group, exacerbated inequality at different levels of technology literacy, breach of basic human rights including voting rights and freedom

Korea

The MSIT, responsible for strategizing and implementing science, technology, and digital innovation in Korean society, has been actively involved in enhancing the social acceptability of AI technology by developing and shaping the bedrock for AI policies, namely the **Digital Bill of Rights** and the **National AI Ethical Standards**.^{liv} In December 2020, MSIT launched the first National AI Ethical Standards, drawing from recommendations provided by organizations such as the OECD and EU. Of the ten standards laid out by the MSIT, the first focuses on the protection of human rights and the third on respect for diversity.

In May 2020, MSIT initiated the **AI Legislation and Regulation Committee**, charged with pinpointing areas for enhancement in the current regulatory framework to more effectively address AI's impact on Korean society. In the final roadmap, unveiled in December 2020, the Committee identified the need for a legal foundation for digital inclusion policies that could strengthen citizens' capabilities and narrow gaps so that the benefits of AI could be distributed more universally.^{lv} To address this need, the roadmap proposed the enactment of the **Digital Inclusion Act**.

Although the legislative process for the Digital Inclusion Act was paused in recent years, the government reaffirmed its support for the Act during the presentation of the Digital Bill of Rights in September 2023, clarifying that the former will be developed based on the latter.^{lvi}

In this regard, the MSIT has introduced **accessibility standards for websites, mobile apps, and kiosks** and promoted the **construction of a kiosk UI platform**, where easy user interface (UI) design guidelines and relevant resources are compiled together. In June 2023, MSIT, in collaboration with NIA and the ICT Accessibility Standardization Forum, hosted the **Digital Accessibility Conference**.^{lvii} Digital platform companies, as well as academia and civil society representatives were invited to share best cases to enhance digital accessibility and discuss design guidelines for digital devices and services that accommodate diverse needs, including the elderly and people with disabilities.

In addition to pushing for digital accessibility in the private sector, so that the future brought by digital and AI transformation is inclusive and diverse, the Korean government is also taking measures to improve the accessibility of its public services. As part of its **Digital Government Masterplan 2021-2025**, the Korean government is dedicated to realizing the mission of implementing intelligent public services aided by AI-enabled virtual assistants.^{lviii} These virtual assistants aim to alleviate the challenges citizens face when accessing digital public services, including difficulty in finding certain benefits or services and other inconveniences related to navigating the digital offerings. Future plans include integrating AI-enabled virtual speakers for vocal guidance, further improving the accessibility to digital public services.

Since 2019, NIA has spearheaded the **Digital Inclusion Forum**, a public-private platform dedicated to addressing digital marginalization at a societal level. During the Steering Committee meeting held in November 2023, the Forum established a standalone subdivision for inclusive policy. This new subdivision joins the existing ones, each dedicated to capacity development, information accessibility, and counter-responses, respectively, enabling mid- to long-term discussions on digital inclusion.^{lix}

In summary, domestic developments in Korea acknowledge the significance of accessibility as an important facet of human rights and freedom in a digital society. However, there appears to be a gap in translating this recognition into concrete processes and practices that bolster AI accessibility within the AI governance framework. As of now, there are preliminary discussions focusing on incorporating accessible and inclusive design in AI development and deployment, improving access to extensive data for AI experimentation and innovation, and enhancing accessibility to public educational resources about and through AI.

"At the current juncture, as we enter the era of heightened digitalization, digital accessibility is becoming a universal right. To ensure the well-being of all citizens, including marginalized groups, in future societies, it is not only essential to enhance their capability to effectively utilize digital tools but also imperative for technology itself to evolve with considerations for diversity and inclusivity. Companies should adopt an inclusive perspective in technology development, implementation, and device design. Simultaneously, the government should strive to facilitate such transitions within companies.

The National Information Society Agency (NIA) supports the enhancement of basic digital skills and practical education for all citizens through the 'Digital Learning Hub.' Moreover, we contribute to the proliferation of a healthy digital information culture and the creation of an ecosystem for digital social innovation. These activities span various fields, including the development and implementation of essential digital inclusion policies, ensuring access to and utilization of intelligent information services, and fostering and nurturing digital talent."

- Choi Moon-sil, Director, National Information Society Agency (NIA) Department of Digital Inclusion

Global developments

The emphasis on AI accessibility is apparent in the global discussions of AI governance, particularly in support of and warranting respect for human rights. Notably, the OECD AI Principles include provisions covering inclusive growth, human-centered values, and fairness.^{lx} Ensuring digital accessibility through human-centered design, diversity and inclusion, and educational programs is a natural consequence of these principles. In fact, OECD's first values-based principle focusing on inclusive growth highlights the importance of engaging in policy action surrounding AI that includes everyone and avoids negative impacts on vulnerable groups.

The National Institute of Standards and Technology (NIST) produced the AI Risk Management Framework (AI RMF) in 2023, which underscores the importance of fairness in creating trustworthy AI systems.^{lxi} One concern raised by the AI RMF is surrounding the inaccessibility of AI systems for individuals with disabilities or for those affected by the digital divide. In

addressing this concern, the AI RMF recommends prioritizing "workforce diversity, equity, inclusion, and accessibility processes" in mapping, measuring, and managing of AI risks.

In the US White House Executive Order announced in October 2023, accessibility is referenced only within the context of providing "workforce training and development that is accessible to all."^{lxii} A broader interpretation might suggest that considerations for AI accessibility may be encompassed within the principle of Advancing equity and civil rights that prohibits discrimination, bias, and other abuses.

Moreover, participants in the Bletchley Declaration, a key outcome from the UK AI Safety Summit signed by 28 countries, agreed that AI should be used in an inclusive manner and noted that in order for AI to be inclusive it must be accessible.^{lxiii}

Overall, recent global developments on AI governance showcase a human rights-centered approach, which inherently promotes accessibility in AI systems and products. As such, there remains plenty of scope

Governments worldwide are leveraging AI to keep pace with public service innovation. With the help of AI, governments are becoming increasingly accessible, available, and attuned to the needs of their citizens in the digital era. These nations stand at the forefront of ushering in inclusive AI in the public sector, and in the heart of their transformative journey, Microsoft remains a dedicated supporter, assisting them in empowering every person within their societies and beyond.

Australia	Canada	India
Australian Public Service will be one of the first governments in the world to trial generative AI services, powered by Microsoft 365 Copilot. The trial will run from January to June 2024. Prime Minister Antony Albanese stated that "by strengthening our partnership with Microsoft, we are charting a course for the future of public service—one where generative AI is used responsibly to enhance the work of the APS in delivering for Australians without compromising on safety." ¹	Microsoft's AI-powered Translator now provides services in Inuktitut, the native language of over 70% of Nunavut's Aboriginal population. Through collaboration with the Government of Nunavut and local community groups, Microsoft has contributed to breaking language barriers in Canadian society and preserving indigenous languages, which are fundamental elements of Canadian culture and heritage. ²	Powered by language models from government-backed AI4Bharat and reasoning models from Microsoft Azure OpenAI Service, the Jugalbandi chatbot assists citizens overcome language barriers and access government programs in their native languages. ³ The chatbot has expanded to cover 10 of India's 22 official languages and 171 government programs, offering personalized information retrieval. The Indian government will keep rolling out language solutions at scale as digital public goods, enabling smoother interactions between citizens and institutions within Indian society.

¹ Prime Minister of Australia (2023). Australian Government collaboration with Microsoft on artificial intelligence. Available at: <https://www.pm.gov.au/media/australian-government-collaboration-microsoft-artificial-intelligence>
² Microsoft (2022). "Introducing Inuinnaqtun and Romanized Inuktitut!" Available at: <https://www.microsoft.com/en-us/translator/blog/2022/02/01/introducing-inuinnaqtun-and-romanized-inuktitut/>
³ Microsoft (2023). "With help from next-generation AI, Indian villagers gain easier access to government services." Available at: <https://news.microsoft.com/source/asia/features/with-help-from-next-generation-ai-indian-villagers-gain-easier-access-to-government-services/>

for improvement to consider the implications of AI development and deployment for groups with disabilities, based on which measures to address their existing needs in the current digital society, as well as new challenges arising from AI expansion, should be suggested.

Private Sector

For Microsoft, accessibility is not a mere buzzword or a vague principle; it is an inherent aspect of doing business. Recognizing the transformative potential of technologies, especially AI, on the lives of people with disabilities, Microsoft created a dedicated accessibility research team. This team focuses on developing innovative solutions to bridge the disability divide prevalent in society. Since 2018, Microsoft has been committed to empowering people with disabilities through the AI for Accessibility program.^{lxiv} This initiative aimed to expedite the development of more inclusive AI solutions and normalize the practice of developing and deploying AI systems with accessibility in mind within the AI ecosystem. This is exemplified through Microsoft's various accessibility-enhancing tools including Accessibility Insights, a developer tool for UI accessibility testing and remediation, designed to reduce unintended barriers for users with disabilities. Over the course of 5 years, AI for Accessibility invested USD 25 million, and in 2021, Microsoft launched a new accessibility commitment that focuses on technology, the workforce, and the workplace.^{lxv}

Amidst the swift digitalization of information, Microsoft has played a crucial role in safeguarding basic human rights, specifically voting rights for people with disabilities, by leveraging technology to enhance digital accessibility. In collaboration with the American Association of People with Disabilities (AAPD), Microsoft introduced the Disability Voting Index in 2022.^{lxvi} This comprehensive tool compiles information about accessible voting options in all 50 states in the US, organizing it into a searchable database for convenient access. The Disability Voting Index stands as a tangible outcome of Microsoft's commitment to closely engaging with disability communities, actively addressing their unique challenges in the digital era through technological solutions.

Microsoft's dedication to AI accessibility extends globally, including Korea. In 2020, Microsoft collaborated with SK Telecom and the Korean Employment Agency for People with Disabilities to conduct an online seminar on the topic of AI for Accessibility. During the seminar, Microsoft introduced its various AI-enabled assistive technologies that people with disabilities can utilize to enhance their work, life, and human interactions and sought cross-sector cooperation to foster an inclusive society that recognizes diversity.^{lxvii}

In the same year, Microsoft launched the Microsoft Enabler Program in Korea, among other countries, which has since expanded to 9 countries and has helped over 350 individuals with disabilities in securing job opportunities.^{lxviii} Microsoft supported its local partners in instilling a

culture of diversity and inclusion in their organization through 45 hours of training focused on designing an inclusive hiring process and the use of assistive technologies implemented through agile AI. This support has been possible and proven to be effective due to Microsoft's years of investment into accessibility research and thought leadership, based on internal trials and insights.

For Microsoft, accessibility is a responsibility, and an opportunity. AI plays a significant role in addressing the social challenge of integrating people with disabilities into society.^{lxix} Simultaneously, Microsoft also sees accessibility as a business opportunity to embrace a previously marginalized group as its avid users and appeal to value-driven millennials. Thus, Microsoft has long incorporated considerations of accessibility into its products and processes, accruing extensive experience and expertise in operationalizing AI accessibility for a more inclusive digital society. This serves as an indispensable pool of knowledge for policymakers worldwide.

Principle 2

Guarantee of fair access and equitable opportunities in the digital

Competition and opportunities for innovation should be equitable in the digital society, ensuring that the benefits of digital innovation are enjoyed by the entire community. Principle 2 identifies five provisions as below to ensure fair access and opportunities.^{lxx}

<p>Promotion of fair competition</p> 	<p>To foster a fair competitive environment in the digital economy, appropriate measures should be taken to address the harms caused by information and technology monopolies, as well as issues of unfairness in algorithms.</p>
<p>Protection of digital properties and assets</p> 	<p>Digital assets formed through individual investment and efforts should receive fair protection, and contracts related to these transactions should be fair, ensuring individuals the freedom to enter into them.</p>
<p>Enhancement of digital literacy</p> 	<p>The digital divide must be addressed to ensure equal opportunities for the development and use of digital technology, and educational opportunities for enhancing digital literacy should be provided.</p>
<p>Guarantee of data access</p> 	<p>Open data access should be promoted, and especially in the case of public data, equal opportunities for access and use must be ensured fairly, with necessary measures taken to universally expand the right to its usage.</p>
<p>Enabling social safety nets</p> 	<p>Benefits of digital innovation should be shared by the community, with necessary measures taken to enhance social safety nets to address economic and social inequalities resulting from digital advancement.</p>

Given the shrinking population in Korean society, attributed to the world's lowest birth rate, enhancing the productive capacity of the future workforce has and will continue to become crucial for achieving mutual prosperity. In a digital society, productive capacity can be read as a direct function of the population's level of **digital literacy**, with a special emphasis on AI literacy. The table below provides examples of how an AI governance framework may aim at improving digital literacy.

Provision	Contextualization in AI governance	Underlying societal concerns
<p>Enhancement of digital literacy</p> <ul style="list-style-type: none"> Development of a support scheme for workforce reskilling and transition Integration of digital literacy and AI education into school curriculum Community awareness workshops and campaigns AI ethics teaching materials Active public-private partnerships at provincial and regional levels for AI capacity building Increased research and development in fake news filtering technology 	<ul style="list-style-type: none"> Loss of economic capacity and productivity due to exclusion of non-digital natives, social cost of digitally excluded, increased vulnerability to manipulation and fake news 	

Korea

Recognizing the significance of digital literacy in society, Korea has demonstrated a strong commitment to digital literacy, particularly in the context of AI, through various initiatives. Following the launch of the National AI Ethical Standards, the MSIT initiated the **AI Ethics Policy Forum** in February 2022. This forum brought together experts from academia, industry, and civil society to decode the ethical standards with key outcomes including **teaching materials** and **instructional guides for educators** to effectively implement AI ethics in education.^{lxxi} These guidelines were updated in April 2023, and the AI Ethics Policy Forum plans to further develop its educational resources, transforming them into more specific and accessible content to enhance AI literacy in Korean society.^{lxxii}

The Ministry of Education (MOE)'s efforts complement those of the MSIT as they endeavor to establish **AI ethics in educational environments**. In August 2022, the MOE officially announced the **Ethical Principles for AI in Education**, which serves as a foundational document for the responsible development and use of AI within educational institutions and activities in Korea.^{lxxiii}

The core focus of these principles is to ensure that AI technologies are harnessed in ways that support human growth, uphold human dignity, and nurture robust human relationships within the educational sphere. This document serves as a valuable guide on AI ethics in the context of education for teachers and educational organizations.

Another area of innovation in the public education system that augments students' understanding of and ability to use AI is the **introduction**

and expansion of AI-powered digital textbooks.^{lxxiv} The MOE aims to launch AI digital textbooks initially in subjects like math, English, computer science, and Korean for special education by 2025. Furthermore, by 2028, they plan to expand the use of AI digital textbooks to cover all subjects, including Korean, social studies, and science. The AI integration in this context allows every student to receive personal tutoring and guidance for their learning material, enhancing their learning experience by making guidance readily available and accessible at no additional cost.

As part of its goal to create a robust foundation for the widespread use of technology and promote innovation in AI-powered services, NIA provides **digital competency education on AI and other ICT developments** to the public. The goal is to elevate the digital capabilities of Korean society.^{lxxv} As such, the Department of Digital Inclusion at NIA is committed to the dissemination of digital education for people from all backgrounds and various demographics to ensure that all citizens are not marginalized or excluded from the digital society. In particular, NIA is expanding cooperation with financial institutions and private franchises to provide education tailored to actual consumers such as financial services, transportation services, and e-commerce. Additionally, NIA extends its support to small and medium-sized enterprises (SMEs), public agencies, and startups by providing consulting services and technical education.

In conclusion, efforts with respect to AI and enhancing digital literacy in Korea include integration of AI ethics in education, application of AI to enhance access to quality education, as well as improving the AI-related digital competencies of both Korean citizens and organizations.



Global Developments

Beyond Korea, global policy initiatives place a particular emphasis on labor market transition and skilling. For instance, Principle 2.4 in the OECD AI Principles emphasizes that governments should engage in AI skilling initiatives to ensure a fair transition to an AI-enabled economy.^{lxxvi} This involves policies that support lifelong learning and reskilling, with a focus on equality, diversity, and fairness. Principle 2.4 is reflected in some examples of government initiatives such as those carried out by Australia with its Next-Generation AI Graduates Program, Kenya with its Digital Literacy Program, and Singapore with its Guide to Job Redesign in the Age of AI.^{lxxvii}

Building further on OECD AI Principles, the International Guiding Principles on Artificial Intelligence agreed to by G7 members in October 2023 also include provisions covering digital literacy.^{lxxviii} In Principle 9, organizations are urged to support digital literacy initiatives while prioritizing responsible stewardship of AI system development and adoption.

Moreover, the White House Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence is guided by eight principles of which one focuses on job training and urging organizations to understand the impact of AI on the workforce. The Executive Order directs organizations to develop best practices to mitigate the harmful impact of AI on workers while maximizing its beneficial impacts. The Executive Order also mandates several reports on AI's impact on the workforce including a study on the "labor-market effects of AI" to be produced by the Council of Economic Advisers.^{lxxx}



Participants in the UK-led AI Safety Summit also emphasized the importance of developing AI-related skills and talent.^{lxxxi} They brought up a range of related priorities including providing people with the skills and knowledge to "design, develop, deploy, and use AI" as well as ensuring that people benefit from the new opportunities created by AI.

As such, the strategy of ensuring that people benefit from AI through digital literacy and skilling programs can be observed in global developments related to AI governance.

Private Sector

Tech companies like Microsoft engage in AI skilling initiatives to address the shortage of skilled AI professionals and make the relevant skills accessible to job seekers in an equitable manner. The MS AI School, an initiative by Microsoft Philanthropies, addresses Korea's AI skills needs through comprehensive training courses with incentives and allowances, facilitated by partnerships with government and industry bodies.^{lxxxii} This initiative not only equips job seekers with AI skills but also directly addresses the workforce deficit highlighted by the Software Policy and Research Institute (SPRI).

Personal testimonies illustrate the profound impact of AI skilling. For instance, Lee Seung-Yoon transitioned from a background in cultural heritage conservation to AI and big data after participating in the MS AI School. The intensive six-month training program provided him with a broad understanding of AI, from basic programming to advanced concepts like deep learning. His success story underscores the school's capacity to empower individuals from diverse academic backgrounds to pivot into the burgeoning field of AI.

In addition to its own initiatives, Microsoft supports and welcomes local partner companies' initiatives to address challenges and make society fairer and more equitable. One such example is Jump, an incorporated association supported by Microsoft, that has made significant strides in leveraging AI to bridge the digital skilling gap for 2,300 individuals, including teenagers, young adults, and women with career backgrounds.^{lxxxiii} This initiative goes beyond technical training and encompasses employment mentoring. Eun Cho-rong, Co-CEO of Jump, underscores the commitment to inclusivity in the digital age, ensuring no one is left behind in the digital education sphere. The '2023 Digital Empowerment Project' marks the culmination of a six-month endeavor, offering tailored programs to participants, particularly enabling high school students to gain hands-on experience with AI services and technologies via Microsoft's Azure.

The project took special care to address the underrepresentation of women in technical education. In partnership with AI-specialized company Testworks, AI data labelling training was provided to women from employment-vulnerable groups, alongside employment opportunities. High school students benefited from Microsoft's cloud service Azure,



engaging with advanced AI services like Custom Vision and Machine Learning Designer without the need for coding. They also had the opportunity to acquire AI 900 certifications, a testament to their knowledge in Azure services. The educational experience was enriched with discussions on responsible AI and mentorships by IT professionals.

Private sector involvement in ensuring access to AI skills is important for furthering the cause of AI accessibility at large. Diverse representation in the AI workforce reduces the chance of biased outcomes of AI systems. Meanwhile, these initiatives also serve to reduce skills and employment gaps.

Microsoft also published a report with the Korea Chamber of Commerce and Industry (KCCI) in August 2023 on the economic impact of generative AI in Korea, which quantified the impact of the technology on Korea's productive capacity and provided policy guidance on harnessing its potential.^{lxxiv} According to the research, generative AI can unlock USD 476.3 billion in productive capacity. A digitally literate and capable population is a key ingredient for Korea to adequately capture the benefits of AI adoption. The report also emphasized the importance of recontextualizing basic skills like critical thinking as well as cross-functional skills like coordination that would be impacted by the adoption of generative AI. Insights presented in reports such as this one can be used by other industry players to guide their thinking around reskilling their employees as well as by governments while drafting related policies.

Korea's approach to sustainable AI

Sustainable AI in Korea's governance approach emphasizes the long-term viability and positive societal impact of AI technologies. This includes considerations for environmental sustainability, economic growth, and the overall well-being of communities. Korea recognizes the importance of developing AI technologies that contribute positively to the broader goals of sustainable development. Korea aims to develop and deploy AI in a manner that aligns with broader sustainable development objectives, reinforcing the importance of responsible innovation for the benefit of current and future generations.

Principle 5 of the Digital Bill of Rights exemplifies this approach.

Principle 5 Promotion of the well-being for all humankind

The final principle for fostering a mutually prosperous digital society advocates international solidarity grounded in universal values and mutual trust. This solidarity is crucial for ensuring that digital technology effectively contributes to the enhancement of human welfare by addressing global challenges with technology-driven innovative solutions and the bridging of digital gaps between nations. To collectively harness the advantages of a sustainable digital society, the principle identifies three key areas of focus:^{lxxxv}

Sustainable digital society



Each country should demonstrate solidarity with the international community to minimize digital technologies' adverse impact on the environment, ecology, and the climate and strive to use technology to expand the welfare of the human community.

Resolution of digital disparities between nations



Each country should cooperate with the international community to abate the digital gap between countries and promote international development cooperation in the digital for the common prosperity of global citizens.

Cooperation for international digital norms



Each country should cooperate with the international community to construct a global governance framework for the formation and implementation of universal digital norms and collective responses to digital risks.

Sustainability in the realm of AI governance, as is the case in all other areas, encompasses social, financial, as well as environmental viability and continuity. Global actors are actively participating in discussions to reap the maximum benefits of AI-driven innovation for the expansion of human welfare by resolving global challenges. Essentially, the goal is to harness the benefits of AI while mitigating potential risks and ensuring that the technology contributes positively to the well-being of humanity. Such discussions have mostly taken place at the international level. The table below illustrates how the provision of **Sustainable digital society** in AI governance has unfolded, highlighted potential concerns driving the discourse.

Provision	Contextualization in AI governance	Underlying societal concerns
Sustainable digital society	<ul style="list-style-type: none"> • International declaration • G7 Hiroshima Process • UN High-Level Advisory Body on AI • Multistakeholder engagement • Global Partnership on AI (GPAI) • Establishment of national AI Safety Institutes • A new international organization for AI governance as a standardization body • Promotion of capacity building in developing nations 	<ul style="list-style-type: none"> • Potential trade disputes spiralling from regulatory discrepancies that could hinder international collaboration and innovation, collective pursuit of public good and safety, and exploitation of AI technologies by a party with ill intentions



Korea

Recognizing that the impacts of AI extend beyond national borders, Korea actively participates in global cooperation for AI governance, advocating for the use of AI for sustainability, as well as the sustainable development of AI. A significant initiative in this regard was the MSIT's cooperation with the International Telecommunication Union (ITU) under the **AI for Good program**, which brings together governments around the world, international organizations, academics, and industry to map out a future where AI is used to respond to global problems and achieve the Sustainable Development Goals (SDGs). In February 2023, the MSIT co-hosted a **seminar on AI for improved health and well-being at all ages**.^{lxxxvi} Declaring that the healthcare gap between countries is a threat to the prosperity and sustainability of the international order, as demonstrated by the public health crisis instigated by the global COVID-19 pandemic, the MSIT sought consensus on the critical importance of international cooperation to develop and employ AI to realize the common goal of sustainable development.

In September 2023, NIA inaugurated the **Digital ESG Roundtable**, comprising key digital and tech companies such as Microsoft Korea, SK Telecom, KT, LG UPlus, and NHN.^{lxxxvii} The Roundtable is anticipated to serve as a hub for fostering both public-private and private-private partnerships. Industry stakeholders will use this platform to share and discuss current issues, aiming to improve their ESG practices. Additionally, it will play a role in contributing to the Korean government's research on digital ESG policy and in strengthening the digital ESG ecosystem in Korea and embracing cooperation with global tech companies.

Furthermore, the **Digital Bill of Rights**, which has been formulated with a primary goal of global standardization, calls upon the international community to prioritize sustainability and well-being of all humankind. The decision to announce and propose the five principles in New York emphasizes the international dimension of this initiative. Despite being officially published in September 2023, these principles for constructing a digital society of mutual prosperity including human well-being and sustainability have already gained prominence in global forums, including the UN, the OECD, and the UK. The government has expressed its commitment to "incorporate the contents of the Digital Bill of Rights in the international discourse" and gain a secure footing in discussions on AI, digital norms, and governance.^{lxxxviii}

The **Plan for the Nationwide Mainstreaming of AI** released in September



2023 also articulates how the popularization of AI in everyday life can improve the quality of life of marginalized populations in society, thereby contributing to the overall equity and sustainability of the society.^{lxxx}
^{lxix} Plans include initiatives like AI care robots for seniors living alone, teleconsultation and AI support for small business owners and the application of AI in public healthcare for people with disabilities, among others. By leveraging AI, the MSIT plans to alleviate daily grievances present in Korean society.

"The Digital Bill of Rights emphasizes on the need for technology to minimize its negative impact and harm on the environment, ecology, and climate system, in order to contribute to the creation of a sustainable digital society. It is highlighted that efforts should be directed towards utilizing technology to expand the well-being of the global community. AI enables humanity to address urgent challenges in innovative ways. A digital society should actively develop by leveraging advanced technologies such as AI to seek effective solutions to the climate crisis and enable humanity to fulfill its environmental responsibilities. In this process, close and concrete collaboration between the public and private sectors should be necessary."

- Choi Moon-sil, Director, National Information Society Agency (NIA) Department of Digital Inclusion

Global developments

Global norm-setting efforts in the realm of AI governance are pivotal for promoting the well-being of all humankind. These efforts ensure a standardized and consistent approach to AI regulations across borders, preventing fragmentation and conflicts. The international collaboration facilitated by global norms is essential for addressing the AI-driven challenges, such as enhanced disparity from gaps in digital literacy, sustainability, and climate change. Additionally, these norms provide adaptability to the rapidly evolving AI landscape, ensuring that governance mechanisms remain relevant and effective in the face of technological advancements. Ultimately, global norm-setting efforts serve as a cornerstone for creating a responsible and safe AI landscape that benefits humanity as a whole.

Endeavors to reinforce the emphasis on sustainability manifest consistently in diverse global AI standards and principles. For instance, the OECD AI Principles advocate for inclusive growth, sustainable development and well-being as the first and foremost principle.^{xc} This mandates that trustworthy AI should enable humans, natural environments, and society at large, so that the future is mutually prosperous for all.

The G7 AI Principles mandate the prioritization of addressing the most pressing social challenges, such as climate change, global health, and education, using advanced AI systems.^{xcii} Under this principle, advanced AI systems are to support the progress of the UN SDGs and help enable the broader society to better respond to the world's priority challenges in an innovative manner.

The US White House Executive Order also recognizes the importance of promoting "the safe, responsible, and rights-affirming development



and deployment of AI abroad to solve global challenges.^{xci} One of the key challenges being the advancement of sustainable development, the Executive Order underscores the necessity of international cooperation to support sustainable deployment and use of AI worldwide.

Furthermore, the Bletchley Declaration from the UK AI Safety Summit commits to the transformative use of AI for good and for all. AI systems would be developed and used to “strengthen efforts towards the achievement of the UN SDGs” and promote inclusive development in terms of public services, food security, biodiversity, and climate.^{xcii} To that end, international collaborative efforts would serve as the backbone for establishing a framework that channels AI’s potential to maximize the greater good like environmental sustainability. In fact, the AI Safety Summit itself is a manifestation of such collaboration, bringing together representatives from 28 countries, along with leading tech companies, civil society groups, and academic experts, to deliberate on actions related to AI safety. As Korea prepares for a mini summit in May 2024 as a follow-up to this initiative, Korea maintains its commitment to being an active partner in the global discourse on AI policies to bring about benefits for all.

Private sector

As a fervent advocate for the idea that AI should contribute to the well-being of all humankind, Microsoft’s mission is to empower individuals and organizations worldwide through AI advancements. Recognizing climate change as the defining issue of our generation, Microsoft is deeply committed to sustainability and employs AI to accelerate progress.

Microsoft has committed to achieving carbon negativity, water positivity, and zero waste, all by the year 2030.^{xciiv} Simultaneously, the company is dedicated to safeguarding ecosystems and constructing a “Planetary Computer,” a comprehensive digital platform designed for sustainable decision-making.^{xciv}

Crucially, Microsoft actively shares best practices and lessons with the industry, contributing to the establishment of norms and promoting exemplary practices and processes. Initiatives like the AI for Good Lab and Global Renewables Watch (GRW) provide grants, technology, and data access to empower individuals and organizations in developing innovative AI solutions for environmental challenges and enhancing climate

resilience.^{xci} The AI for Good Lab facilitates access to data from the Planetary Computer and global organizations for local entities, enabling the development of scalable climate solutions. The GRW, a collaboration between Microsoft, Planet Labs PBC, and the Nature Conservancy, uses AI and satellite imagery to assess clean energy transition and track trends publicly. It provides valuable insights into renewable energy capacity at the country level, contributing to the understanding of potential impacts on the landscape over time.

In a notable collaboration, Microsoft partnered with the Sustainable Environment and Ecological Development Society (SEEDS) to develop an AI model predicting the impact of cyclones on vulnerable populations in India.^{xciiv} high-resolution satellite imagery, the model employs advanced data analytics and machine learning to identify houses at the highest risk. This enables SEEDS to precisely target outreach efforts to communities most susceptible to the cyclone’s impact.

Microsoft prioritizes responsible AI development, integrating sustainability into all innovations. Their recent playbook on Accelerating Sustainability with AI outlines the unique capabilities of AI in achieving net-zero goals which includes the ability to measure, predict, and optimize complex systems, acceleration of the development of sustainability solutions, and empowerment of the sustainability workforce.^{xciiv} Moreover, the playbook introduces Microsoft’s five-point action plan from investing in AI, developing digital and data infrastructure, minimizing resource use in AI operations, advancing AI governance, and building workforce capacity to use AI for sustainability for unleashing AI’s power for transformation. All these efforts aim to help ensure that AI is developed and deployed in a direction that can more effectively benefit everyone.

Microsoft’s latest Environmental Sustainability Report highlights their progress towards the 2030 sustainability goals.^{xci} The business saw a 0.5% reduction in overall carbon emissions in 2022 despite growth and Microsoft extended access to clean water and sanitation to nearly 1 million people, in addition to achieving an 82% reuse and recycle rate for all cloud hardware, diverting over 12,000 metric tons of operational waste from landfills and incinerators. Microsoft has intensified investments in more efficient data centers, clean energy, improvements to the Microsoft Cloud for Sustainability and Planetary Computer, and eco-friendly software practices. Through their Climate Innovation Fund, Microsoft has committed over \$700 million to a diverse global portfolio of 50+ investments, encompassing sustainable solutions in energy, industry, and natural systems.

As part of its commitment to global coordination on AI governance, especially in sustainability, Microsoft actively participates in multistakeholder initiatives. Leveraging its experience and expertise, Microsoft stands ready to serve as a valuable partner in advancing climate resilience for all humankind through the accelerated power of AI.

Conclusion

In today's technology-driven world, effective AI governance is crucial, particularly given AI's substantial impact on society and the economy. As a leading global IT nation, Korea must embrace a forward-thinking approach in AI governance, aligning with international best practices while customizing these strategies to cater to the unique needs of its populace.

The Korean Digital Bill of Rights plays a crucial role in shaping Korea's AI regulations and will have a broad impact on subsequent laws, policies, and regulations. It will also influence how regulators and judicial authorities interpret the role of AI in Korean society. This Bill focuses upon five basic principles of freedom, fairness, safety, innovation, and solidarity, all of which are key elements of ensuring the development technology in a mutually prosperous manner for society and industry. The Bill also highlights the importance of including insights from a wide range of stakeholders in the AI industry to ensure their roles and risks are accurately represented.

Emphasizing responsible AI governance is crucial to ensure that AI technologies are developed and deployed responsibly, upholding human rights, and preventing biases in AI algorithms. This responsible approach safeguards transparency and accountability in AI decision-making. Furthermore, the integration of sustainability and inclusivity in AI governance is indispensable. Korea's commitment to these principles ensures that its AI advancements are sustainable, catering to diverse societal needs, and inclusive, providing equitable benefits across all sections of society.

In focusing on Korea's unique position in AI governance, it is important to recognize its technological prowess and potential as a global leader in AI innovation. Korea's technological infrastructure, combined with its forward-thinking policies, positions it well to set benchmarks in AI governance. Adopting a strategy that weaves together responsible, sustainable, and inclusive AI practices will benefit Korea's domestic market and elevate its international standing. Korea's approach can serve as a model for other nations, demonstrating how to balance technological advancement with

responsible considerations and societal welfare. This strategic positioning underscores Korea's role as a pioneer in the responsible and innovative use of AI.

This paper has illustrated the importance of aligning the Korean Digital Bill of Rights with international standards and emphasizes the need for ongoing collaboration between the government and private sector. This strategy can drive innovation within Korea and also enable Korea to be a global leader in AI policy discussions. It is recommended that Korea consistently updates its AI practices to align with global trends, maintaining its position at the forefront of the digital world.

In a broader sense, this paper points out Korea's commitment to a balanced approach in managing AI, aiming to nurture innovation while ensuring safety. This strategy mirrors Microsoft's practices in AI governance, suggesting a shared vision for growth and safety in the AI sector. The successful collaboration between Microsoft and Korean entities such as NIA exemplifies a model partnership, bolstering Korea's aim for a robust and secure AI industry.

In summary, this paper underlines the importance of integrating these strategies into a comprehensive AI governance framework specific to Korea. Such a framework would drive Korea's growth, and also empower its people and organizations in the AI field. The Digital Bill of Rights is a significant opportunity for Korea to lead in AI governance. Adopting international standards and nurturing public-private partnerships enables Korea to formulate AI policies that are effective on both a local and international scale. This strategy is crucial for building a digital society that is responsible, sustainable, and inclusive, and therefore mutually prosperous. By aligning with global leaders, Korea can craft an AI governance model that maximizes AI's benefits and addresses challenges, contributing significantly to Korea's continuous economic and technological growth.

Appendix A

Current status of AI governance in Korea

This section outlines the different governmental entities in Korea and their proactive efforts in shaping AI governance within the country.

Ministry of Science and ICT (MSIT)

The MSIT, responsible with strategizing and implementing science, technology, and digital innovation in Korean society, has been actively involved in enhancing the social acceptability of AI technology by developing and shaping the bedrock for AI policies, namely the Digital Bill of Rights and the National AI Ethical Standards. In December 2020, they launched the first National AI Ethical Standards, drawing from recommendations provided by organizations such as the OECD and EU.^c With the objective of ensuring AI benefits humanity, the standards articulated three core principles around the values of human dignity, public welfare, and purposefulness of technology. Additionally, it outlined ten key requirements:

- 01 Protection of human rights
- 02 Privacy protection
- 03 Respect for diversity
- 04 Harm prohibition
- 05 Publicness
- 06 Solidarity
- 07 Data management
- 08 Responsibility
- 09 Safety
- 10 Transparency

Subsequently in February 2022, the MSIT initiated the AI Ethics Policy Forum, bringing together experts from academia, industry, and civil society to decode these standards into applicable recommendations. Key outcomes of this initiative included self-auditing and development guidelines for companies, along with teaching materials and instructional guides for educators to effectively implement AI ethics education.^d These self-review and development guidelines were updated in April 2023, and the AI Ethics Policy Forum plans to further develop its educational resources into more specific and accessible content for the enhancement of AI literacy in Korean society.^{ei} By December 2023, the second phase of the AI Ethics Policy Forum intends to create sector-specific specialized self-checklists and guidelines for areas where reliability and ethics are emphasized, such as recruitment, public safety, and services using generative AI. These guidelines will offer AI system developers and operators a set of voluntary compliance criteria for testing and certification standards.

In May 2023, the MSIT announced the new Digital New Order, a blueprint that would later evolve into the Digital Bill of Rights. Introduced in September 2023,^{ei} this document outlines the universal values to be pursued in the digital era where AI is extensively used across various industries. While the charter consists of 28 concise, high-level commitments, it is slated to serve as the starting point for defining the rights and responsibilities of citizens, businesses, and the government in resolving AI-associated risks, such as combatting fake news. The MSIT is expected to introduce follow-up regulations in alignment with the principles outlined in the Digital Bill of Rights, such as the AI Act and the Digital Inclusion Act.

In August 2023, the MSIT inaugurated the 4th AI Legislation and Regulation Committee. This committee's primary mission is to revise the AI legislative roadmap initially formulated in December 2020. Their goal is to enhance regulatory frameworks and build a trustworthy foundation for hyper-scale AI. The 4th Committee comprises government officials, industry experts, and academics, distributed across four specialized Subcommittees. These subcommittees address diverse areas, including



regulatory enhancements, trust-building measures, sector-specific considerations, and AI-related copyright matters.^{cv}

According to the Plans for Promotion of AI Ethics and Reliability Assurance announced in October 2023, the MSIT is engaged in the development of a framework for responsible AI governance.^{cv} Major initiatives comprise the regulatory revisions aimed at encouraging voluntary compliance within the private sector, investment in and operation of technology to certify AI trust, the expansion of AI impact assessment and education on ethics, and the promotion of global digital norms, as depicted below.

Regulatory revisions

Enhanced AI safety

- ▶ Criteria guidelines for high-risk AI systems (To be developed in 2024)
- ▶ Guidelines for ensuring trust for high-risk AI systems (To be developed in 2024)

User safety

- ▶ Guidelines for AI watermarks (To be developed in 2024)

AI development and expansion

- ▶ Guidelines for use of publicly available personal information in AI (To be developed by March 2024)
- ▶ Guidelines for AI copyrights (Announced in December 2023)

AI impact assessment and ethics education

Enhanced AI safety

- ▶ Framework for risk identification and management measures for AI services
- ▶ Analysis of social, economic, cultural, and public life impacts of AI products and services (To be continued in 2023)

Customized ethics education

- ▶ Tailored AI ethics education courses for different audience groups and levels
- ▶ Expansion of the target audience (general public in 2023; developers and deployers, etc. in 2024)

Technological assurance of AI reliability

Launch of voluntary AI safety and trustworthiness testing and certification system in 2024

Continuous research and development of next-generation AI technology that can overcome limitations and hyperscale Aaility assurance technology in 2024

AI watermarking system in accordance with global trends (To be promoted in 2024)

International cooperation

Promotion of the Digital Bill of Rights in global forums

Joint AI research with the Middle East, ASEAN, the US and UK

- ▶ AI-based solutions to global challenges
- ▶ AI application in medical, education, energy sectors

Developing pertinent global agenda

- ▶ AI Safety Institute ▶ Global standards
- ▶ International organization for AI governance under UN

Co-host of the mini-AI Safety Summit in May 2024



National Information Society Agency (NIA)

As a government think tank that specializes in digital transformation, the National Information Society Agency (NIA) conducts in-depth research into AI-related legal frameworks to formulate digital policies and resolve societal issues that may arise from the integration of AI into everyday life.^{cvii}

NIA collaborates closely with the MSIT and leads the AI Legislation and Regulation Committee, playing a key role in shaping national discussions on AI governance, aligning with global standards, and ensuring potential risks related to ethics, safety, and trust in AI are comprehensively addressed in the national AI policies.

The agency also facilitates the safe and responsible growth of the AI industry by establishing national infrastructure for large-scale AI learning. NIA operates a national metadata platform, enabling the aggregation and accessibility of valuable national information. This platform enhances the credibility of national metadata, which is essential for large-scale AI learning datasets and helps ensure the trustworthiness of AI technologies.

The agency's broader goal pertaining to AI is to build a robust foundation for the widespread use of the technology, promote innovation in AI-powered services, and foster collaboration between the public and private sectors. As part of this mission, NIA provides digital competency education on AI and other ICT developments to the public, with the aim of enhancing the digital capabilities of Korean society. NIA also extends its support to small and medium-sized enterprises (SMEs), public agencies, and startups by providing consulting and technical education.

Meanwhile, recognizing the importance of AI ethics, NIA is also promoting AI ethics education for youth, AI companies, and developers, and in particular, actively supporting the joint establishment of AI ethics principles, in-house ethics education, and 'Good AI Company' certification to strengthen the AI ethics capabilities of SMEs in the AI industry.

National Assembly of Korea

In February 2023, the Science, ICT, Broadcasting and Communications Committee of the National Assembly passed the draft Act on Promotion of AI Industry and Framework for Establishing Trustworthy AI (AI Act).^{cviii} A consolidation of seven pre-submitted bills pertaining to AI, its title distinctly outlines its dual focus and objectives, which sets it apart from the approach taken by the EU AI Act where the focus on user safety is quite salient. Nevertheless, some voices from the industry have noted certain parallels in the proposed regulatory measures and definitions targeting high-risk AI systems.^{cviii} Notably, the bill includes a clear provision for principles of preferential approval and ex-post regulation for AI technology, often referred to as the 'permit-first-regulate-later principle' within the Korean context. The AI Act sets forth essential principles to guide the safe and responsible development of AI, ultimately aimed at enhancing the lives of Koreans. If passed, the Act would serve as the statutory foundation for overseeing the AI industry in Korea. It would also

give rise to the Basic Plan for AI, an AI committee operating under the Prime Minister's Office to implement the law, along with the establishment of a National AI Centre under NIA to promote the industry. As of October 2023, the bill is still pending review by the Standing Committee.^{cxix}

Separately, the Artificial Intelligence Responsibility and Regulation bill was recently introduced to the National Assembly.^{cx} This bill more closely mirrors the scope and approach of the EU's AI system classification by anticipated risks and the prohibition of certain types of AI systems. It is yet to be determined if this signifies a shift in the National Assembly's overall perspective on AI governance and how this bill will advance with the AI Act ahead in the legislative process. There is also the possibility that parts of the whole of this bill may be incorporated into the AI Act and subsequently repealed, increasing the regulatory stringency of the AI Act.

In addition to the introduction of new systems for AI governance, the National Assembly has been integrating considerations of AI responsibility, fairness, trustworthiness, and transparency into Korea's existing regulatory frameworks through legislative amendments. The most recent amendment, which took place in March 2023 and will take effect on March 15, 2024, is to the Personal Information Protection Act (PIPA). This amendment grants data subjects the right to refuse decisions made by completely automated systems for processing personal information if these decisions significantly impact their rights or obligations. Data subjects are also entitled to request explanations for such decisions.^{cxii}

Other bills for amendment encompass:

Relevant legislation	Date submitted	Detail
Personal Information Protection Act	February 2023	This proposal seeks to empower the Personal Information Protection Commission to be able to request the submission of algorithms if a data breach occurs in an AI company's algorithm. ^{cxii}
Fair Hiring Process Act	March 2023	The proposal imposes a prior notification obligation on companies when using AI in their hiring processes. ^{cxiii}
Content Industry Promotion Act	May 2023	The bill proposes labelling content created using AI technology as such. ^{cxiv}
Copyright Act	June 2023	The amendment aims to establish explicit criteria for copyrighted materials in the context of automated information processing. It also seeks to clarify the boundaries of AI technology application and copyright infringement for future use of AI-generated works. ^{cxv}
Public Election Act	June 2023	The amended bill prohibits the transmission of false or AI-manipulated information for campaign purposes and places restrictions on reporting or providing commentary on such manipulated data or information. ^{cxvi}
Information and Communications Network Act	February 2023	This proposal imposes reporting obligations on those providing AI-based recommendations services to the Korea Communications Commission. ^{cxvii}
Information and Communications Network Act	June 2023	The bill intends to amend the definition of an information and communication service provider to include entities that provide information or mediates information using AI technology. ^{cxviii}

All of these proposed amendments remain under subcommittee review as of December 2023.

Personal Information Protection Commission (PIPC)

In addition to coordinating the process of amending PIPA, the Personal Information Protection Commission (PIPC) actively oversees data privacy issues in AI governance. On June 23, 2023, the PIPC hosted the AI and Data Privacy International Conference.^{cxix} During the conference, they discussed the direction of AI regulation with data protection authorities from major countries, such as the UK, Germany, and Japan, along with international organizations like the EU and OECD, and AI experts. This conference spotlighted PIPC's strategic policy direction in managing the dynamic realm of AI and data privacy regulations, underscoring a commitment to harmonizing innovation with the safeguarding of individual rights and interests. The PIPIC is transitioning from a rule-based to a principle-based regulation, advocating for adaptable application of core principles to diverse real-world scenarios. They are also advocating for risk-based, differential regulatory measures and emphasizing the importance of public-private partnerships to establish robust, practical standards.

These approaches have been consolidated and published as the Policy Directions for the Safe Use of Personal Information in the AI Era. This policy outlines the PIPC's interpretation of the current PIPA within the AI environment. The AI Privacy Team will be established to directly interact with AI system and service providers to reduce uncertainty in legal interpretation and compliance. Furthermore, the policy provides clear guidelines and standards for handling personal information throughout the AI system lifecycle, from inception to data collecting, AI learning, and service provision. Following the blueprint for jointly designing regulatory systems with the government and the private sector in specific areas, area-specific guidelines are slated to be developed through collaboration between the government, industry actors, academia, legal experts, and civil society, facilitated by the creation of an AI Privacy Public-Private Policy Council.

Korea Communications Commission (KCC)

The Korea Communications Commission (KCC) has been instrumental in sculping Korea's AI policy, with a strong focus on protecting user rights in our increasingly AI-dominated society. In November 2019, in collaboration with the Korea Information Society Development Institute (KISDI), the KCC unveiled the 'Principles for a User-Centered Intelligent Information Society,' marking Korea's first foray into AI ethics guidelines. These guidelines are designed to create a safer environment for intelligent information services, considering the challenges posed by emerging technologies like AI, big data, and the Internet of Things in broadcasting and communication sectors.^{cxxi} The principles emphasize human-centric services, transparency, and explainability, responsibility, safety, non-discrimination, user participation, privacy, and data governance, highlighting the crucial need to protect user rights and maintain human dignity amidst the rapid integration of AI technologies in everyday life.

In line with this commitment, the KCC has been hosting the annual International Conference on Ethics of the Intelligent Information since 2019, providing a platform for stakeholders to deliberate on the societal

and cultural implications of AI's proliferation.^{cxxii} The KCC aims to harmonize its principles and policies with international standards, fostering global consensus through ongoing forums.

The KCC continues to engage with a diverse group of stakeholders, including users, businesses, and experts, to refine and evolve its AI principles in response to the dynamic AI technological landscape. This led to the June 2021 publication of the 'Basic Principles for Ethics in AI-Based Media Recommendation Services,' with further explanatory documents released in April 2022.^{cxxiii} These principles, reiterating the importance of transparency, fairness, and accountability, advise AI-based recommendation service platforms to prioritize transparent systems and enhance user control over AI-driven content. The guidelines emphasize the development of internal protocols and self-assessment mechanisms to mitigate adverse AI system impacts and address user complaints effectively. The explanatory guide offers detailed guidance to digital media platform operators regarding AI transparency and explainability, balancing disclosure requirements with the protection of trade secrets, and stipulating clear disclosure norms for personal data usage.

Furthermore, the KCC has established a public-private consultative council for AI ethics, partnering with KISDI.^{cxxiv} This council engages in discussions on a range of timely topics concerning user protection in the AI era, such as responsible use of the metaverse, ethics in the age of algorithmic journalism, and generative AI chatbot services.^{cxxv} These topics reflect the emerging user protection challenges in the dynamic landscape of AI technologies, and the KCC's ongoing commitment to addressing them.

Ministry of Education (MOE)

In August 2022, the Ministry of Education (MOE) declared the 'Ethical Principles for AI in Education,' a cornerstone document guiding the ethical development and use of AI in Korea's educational institutions and activities.^{cxxvi}

These principles focus on harnessing AI to foster human growth, uphold dignity, and strengthen human relationships in the educational domain. Developed through comprehensive public consultations, expert meetings, and the incorporation of international insights, these principles represent a commitment to responsibly integrating AI in educational contexts.

These principles are of a voluntary and self-regulatory nature and comprise ten specific guidelines designed to ensure the stable establishment of AI ethics in educational environments. They encompass various provisions, including promoting equal opportunities and fairness in education, contributing to the broader social welfare, ensuring transparency and comprehensibility in data handling, and safeguarding the well-being of all educational stakeholders, among others. The foundational principle that underpins these ethical standards is that AI in education should be leveraged to support human growth, given that the very essence of education is to facilitate human growth.

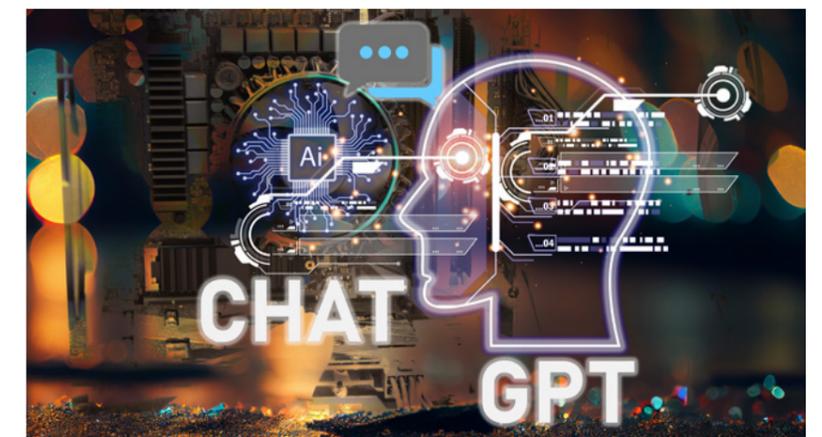
The MOE is proactively planning to integrate these standards across various facets of the education sector. These will be integrated into AI ethics education, materials for teacher training, and guidelines for educational technology (EdTech) companies. The primary goal is to maximize the benefits of AI while ensuring its ethical application in practice. To maintain the relevance and applicability of these principles, the MOE plans to conduct periodic reviews. This will help them adapt to the ever-evolving landscape of AI technology and its usage, keeping the ethical framework up to date.

The MOE is enthusiastic about the integration of AI digital textbooks in the education sector,^{cxxvii} a concept recently legitimized by the amendment to the Regulations regarding Textbooks for Educational Use at the State Council.^{cxxviii} The MOE plans to introduce AI digital textbooks initially in subjects like mathematics, English, computer science, and Korean for special education by 2025. By 2028, the initiative will expand to encompass all subjects, including Korean, social studies, and science. Supporting this initiative, the MOE has partnered with the MSIT via a memorandum of understanding (MoU) to secure technological assistance for these educational innovations they envision.^{cxxix}

National Intelligent Service (NIS)

The rapid adoption of generative AI, particularly ChatGPT, was taking place across various sectors, including public institutions, but there was a lack of clear security measures at the government level. This prompted the National Intelligent Service (NIS) to draft and release the Guideline for Safe Use of Generative AI Including ChatGPT in June 2023, a framework for government agencies, educational institutions, and the general-public to navigate the use of these technologies securely.^{cxxx}

Developed through public-private consultations involving academia, industry, and research institutions, this security guideline provides detailed, step-by-step instructions covering the entire process of using generative AI technologies, from accessing services to drafting prompts and utilizing results. Key principles include prohibiting the input of sensitive information, such as private or confidential data, revalidating the accuracy, ethics, and suitability of generated content, filtering any violations of intellectual property or copyright laws when using generated content, verifying the safety of interconnections and extended programs, and strengthening security settings for login accounts.



Appendix B

Recent global developments in AI governance



Global efforts to provide model references for AI governance are on the rise, as demonstrated by recent announcements by the EU, G7, and the US.

The last few years saw a rise of global efforts to collaborate on AI governance. Most notably, the US, the EU, and G7 have consecutively introduced principles and frameworks to aid the development of safe, secure, and trustworthy AI.

OECD AI Principles

The Organization for Economic Cooperation and Development (OECD) published the Principles on Artificial Intelligence (AI Principles) in 2019. The OECD had been undertaking research on the economic and social impacts of AI technologies and applications since 2016. Having found that there is a policy need for AI governance, the OECD Committee on Digital Economy Policy (CDEP) agreed to develop Council Recommendation “to promote a human-centric approach to trustworthy AI.”^{cxxxix}

Against this background, an expert group, comprising of representatives from government, industry, civil society, trade unions, amongst others was formed to assist the development of the Council Recommendation.^{cxxxix} This was subsequently endorsed and adopted as the OECD AI Principles.

The OECD AI Principles’ overarching goal is to provide a framework for OECD members and non-members to “promote and implement policies for responsible stewardship of trustworthy AI.”^{cxxxix}

The OECD AI Principles comprise five value-based principles, emphasizing key themes like transparency, explainability, robustness, security, safety, and accountability. The OECD AI Principles also entail national policy and international cooperation recommendations for the operationalization of the value-based principles.

Nationally, the OECD AI Principles encourage governments to consider investing in AI research, fostering an ecosystem for trustworthy AI, creating a policy environment that supports the development and deployment of trustworthy AI, and developing plans to support workforce transitions as AI advances. Internationally, the OECD AI Principles recommends international cooperation on trustworthy AI in the form of knowledge sharing, and standards development.

Since the introduction of the OECD AI Principles in 2019, over 71 jurisdictions have developed national AI policies. Countries are also in the process of developing responsible frameworks and principles in accordance with the OECD AI Principles, such as Korea, Japan and India.^{cxxxiv} Additionally, the OECD noted countries’ efforts in translating value-based principles into concrete and operational policy options.^{cxxxv} Most importantly, the OECD AI Principles paved the way for further global cooperation. One notable example is the collaboration between the G7, OECD and GPAI, which resulted in the G7 Guiding principles on AI and the code of conduct for advanced AI systems.^{cxxxvi}

G7 Guiding Principles for Developing AI

The G7 (Canada, France, Germany, Italy, Japan, the UK, US, and the EU) reached an agreement on the International Guiding Principles on Artificial Intelligence (Principles) in October 2023. The Guiding Principles is the output of the “Hiroshima Process,” in which members of the G7 agreed to advance discussions on AI governance and interoperability for trustworthy AI that aligns with the group’s shared value and builds further on the 2019 OECD AI Principles.^{cxxxvii}

Against this background, the eleven Principles are introduced with a clear objective to promote safe, secure, and trustworthy AI.^{cxxxviii} Though it is voluntary and non-binding in nature, the G7 intends for the Principles

to provide guidance to governments and organizations towards the responsible use of AI.^{cxix} To respond to rapid technological advances, the G7 will update and review the Principles as necessary.^{cxl}

The International Guiding Principles on Artificial Intelligence has eleven principles.

- 01 Adopting measures to address risks throughout the development of advanced AI systems
- 02 Addressing vulnerabilities after deployment on the market
- 03 Publicly reporting on advanced AI systems to ensure transparency
- 04 Reporting of incidents among organizations developing advanced AI systems
- 05 Adopting AI governance and risk management policies grounded in a risk-based approach
- 06 Investing in security controls across the AI lifecycle
- 07 Adopting measures that users to identify AI-generated content
- 08 Prioritizing research and investment in risk mitigating measures
- 09 Prioritizing the development of advanced AI systems for global benefit
- 10 Adopting international technical standards
- 11 Implementing data input controls and protections for personal data and intellectual property

On the whole, the Principles cover advanced AI systems, such as foundational models and generative AI. It outlines a non-exhaustive list of actions which can be undertaken in the use, design, development and deployment of advanced AI systems.

At its core, the Principles emphasize applying a risk-based approach throughout the AI lifecycle and ensure that measures such as testing and mitigation techniques, are in place to identify and mitigate risks. Additionally, risk management policies are encouraged to be developed for advanced AI systems.^{cxli}

Given the widespread use of AI, the Principles also stresses the need for AI actors to develop and deploy content authentication and provenance mechanisms, such as watermarking, for users to identify AI-generated content.^{cxlii}

Apart from internal controls and policies, AI actors and organizations are also encouraged to engage in information sharing and exchange to increase accountability. Such include incident reporting, and publicly stating advanced AI systems' capabilities, limitations, appropriate use, and patterns of misuse.^{cxliii}

Keeping in mind of the global nature of the Principles, the G7 called for AI actors to prioritize research to mitigate social, safety and security risks of AI systems, investments in systems which address global challenges, and development of international technical standards.^{cxliv}

The White House Executive Order on the Safe, Secure, and Trustworthy Artificial Intelligence

The Principles serve as a first global initiative to address regulatory challenges introduced by advanced AI systems. Though voluntary in nature, the Principles provide an important reference for AI actors and national governments to consider the risks of advanced AI systems, and the actions which can be taken to address and mitigate such risks. The Principles also pave the way for further global alignment in AI governance by harmonizing with other regulations and principles, such as the EU AI Act and the UK AI white Paper.^{cxlv}

Earlier this year, the US government secured a voluntary commitment from a dozen leading AI companies, including Microsoft, to undertake steps in creating safe, secure and trustworthy AI systems. Building on this momentum, President Joe Biden signed an Executive Order on Artificial Intelligence (AI) in October 2023.

Continuing from the voluntary commitment, the overarching goal of the White House Executive Order is to promote safe, secure and trustworthy development and use of AI. Apart from governing AI companies, the White House Executive Order is applied to organizations across all sectors; and covers a broad range of AI systems including generative AI, and any machine-based system that makes predictions, recommendations, or decisions.^{cxlvi}

The White House Executive Order laid out 8 principles in areas such as AI standards, privacy, equity and civil rights, consumer and worker protection, innovation and competition, and AI governance leadership. There are several principles that are relevant to global AI governance.

As AI continues to advance, the White House Executive Order seeks to set new standards for AI Safety and Security to guard against the potential risks of AI systems. Such requirements include developing standards, tools, and tests to ensure the safety, security, and trustworthiness of AI; mandating developers to share safety test results and critical information with the government; and formulating guidance on content authentication and watermarking to identify AI-generated content.^{cxlvii}



Additionally, the White House Executive Order emphasizes the need to promote responsible innovation, competition, and collaboration through investing in AI research and promoting a competitive AI system.^{cxliii}

Given the global impact of AI, the White House Executive Order directs the government to expand multistakeholder engagements on AI and participate in the development and implementation of AI standards with international partners. Other than collaboration, the White House Executive Order sees the deployment of AI to solve global challenges, mitigating dangers to critical infrastructure, as a priority.^{cxlix}

The White House Executive Order has a significant impact on global AI governance. Though the Executive Order is domestic, it made explicit references to direct the federal government to collaborate with international partners on AI standards. Soon after the issuance of the Executive Order, the US signed the Bletchley Declaration with the EU and 27 other countries to examine and address the impact and risks of AI systems.^{cl} The National Institute of Standards and Technology (NIST) under the Department of Commerce further announced the establishment of the U.S. AI Safety Institute Consortium, which is intended to “equip and power collaborative establishment” to promote the use of safe and trustworthy AI globally.^{cli}

US NIST AI Risk Management Framework (AI RMF)

Before the release of the Executive Order, NIST has been contributing to research, standards, and data to aid the development of AI-related frameworks and policies.^{clii}

Most notably, in January 2023, NIST produced the AI Risk Management Framework (AI RMF). The framework was developed in close collaboration with private and public sectors, and seeks to help “individuals, organizations and society” understand and manage risks associated with AI.^{cliii}

The Framework provides an overview of three types of AI Harms and outlines seven characteristics of trustworthy AI systems. It further offers four interconnected functions which AI actors can adopt across the AI lifecycle to manage the harms effectively.^{cliv}



In terms of AI Harm, the framework categories harms into three broad types, namely, harm to people; harm to an organization; and harm to ecosystem.^{clv} To avoid these harms, the framework pinpoints seven characteristics which AI developers can consider when designing, developing, using, and evaluating AI products, services, and systems. According to the framework, trustworthy AI systems should be valid and reliable, safe, secure, and resilient, accountable, and transparent, explainable, and interpretable, privacy-enhanced, and fair with harmful bias managed.^{clvi}

Designing AI systems according to these characteristics would require the incorporation of various activities and functions. In the framework, it is suggested that AI actors should begin by instilling the Map function, which aims at identifying AI risks by gathering perspectives from different stakeholders.^{clvii} Once that understanding is in place, there should be a Measure function to routinely analyze, monitor, assess and benchmark AI risks.^{clviii} Regular reviews of AI risks should allow actors to adhere to the Manage function, where resources can be allocated to appropriately manage and mitigate AI risks.^{clix} Finally, relevant structures, systems and processes should be put in place to ensure the continuation and governance of the above functions.^{clx}

The AI RMF provides an important foundation for AI governance. Though AI RMF is voluntary, sector and use-case agnostic in nature, NIST issued a series of companion materials, such as the NIST AI RMF Playbook, AI RMF Roadmap to help organizations to implement the approaches in the framework.

After the release of the AI RMF, NIST also illustrated its alignment with other international standards, guidelines, and standards to ensure harmonization. For instance, NIST released a Crosswalk document to demonstrate how the AI RMF’s trustworthiness characteristics relate to the OECD Recommendations on AI, Blueprint for an AI Bill of Rights, Executive Order on Promoting the Use of Trustworthy AI in the Federal Government, and the Proposed EU AI Act.^{clxi}

UK AI Safety Summit

The UK AI Safety Summit, which took place in early November, was the first high level summit convened by the UK Government on the safe development of AI. The Summit invited over 150 representatives from governments, industry, and civil society to discuss AI risks and opportunities, as well as global cooperation options to advance AI safety.^{clxii}

The key objectives set forth in the Summit are to establish a shared understanding of the risks and opportunities of frontier AI and deliberate on the measures that organizations should take to increase the safety of frontier AI.^{clxiii} Simultaneously, the Summit is intended to create a forward process for international collaboration, specifically on AI safety research, and supporting national and international frameworks.^{clxiv}

The key objectives above are fulfilled by the Bletchley Declaration. Signed by 28 countries, the Declaration reached an initial agreement on frontier AI, and acknowledges the opportunities and risks associated with it.^{clxv} The Declaration also expresses the need for national and international cooperation on risk identification and risk mitigation.^{clxvi}

In terms of risk identification, the Declaration expressed the need to build a “shared scientific and evidence-based understanding” of AI safety risks, and a global approach to investigating the impact of these risks in societies. In terms of risk mitigation, the Declaration calls for countries to adopt risk-based policies and establish legal and regulatory frameworks based on national circumstances. This framework could include calls for greater transparency from frontier AI private actors and introducing tools for safety testing.^{clxvii}

The two-day Summit created a space for participants from a wide range of sectors to discuss issues pertinent to global AI governance. In particular, the Summit raises awareness of the need for continuous discussions on a shared understanding of frontier AI, and the urgency to address short-term AI risks brought about not only by frontier AI, but other types of AI. Further, it cemented the recognition of common principles and codes, including the G7 Guiding Principles.^{clxviii}

EU AI Act

The Proposal for a Regulation Laying Down Harmonized Rules for Artificial Intelligence, also known as the EU AI Act, was the first comprehensive legal framework focusing on AI governance in the EU. Put forward by the European Commission, the overarching objective of the EU AI Act is to “ensure the proper functioning of the European single market by creating conditions for the development and use of trustworthy AI systems in the European Union.”^{clxix}

The EU AI Act outlines a series of specific goals. This proposed legislation strives to guarantee the safety of AI systems placed on the market while ensuring compliance with existing EU regulations. It also seeks to harmonize AI regulations to prevent market fragmentation. Additionally, the EU AI Act aims to offer legal clarity, promoting an environment conducive to investment and innovation in AI.^{clxx}

To ensure accountability, the EU AI Act is set to apply to various actors in the AI ecosystem. This includes providers and users within the EU, or in a third country with a view to “placing AI systems in the EU market or putting them into service in the EU.”^{clxxi} However, exemptions are made to AI systems “developed or used exclusively for military purposes,” “public authorities in a third country,” and “international organizations or authorities using AI systems in the framework of international agreements for law enforcement and judicial cooperation.”^{clxxii}

The cornerstone of the EU AI Act is the risk-based approach towards AI

governance. Under the Act, AI systems are classified based on the level of risks they pose to the “health, safety, and fundamental rights of a person.”^{clxxiii} The risks are divided into four broad categories: Unacceptable, High, Limited, and Minimal. Each risk category comes with a corresponding set of requirements and obligations for developing, using, and placing the AI systems on the EU market.^{clxxiv}

Broadly, Unacceptable AI systems, which are prohibited under the EU AI Act, refer to those that employ subliminal, manipulative techniques, and exploit specific groups.^{clxxv} Examples of AI systems in this category include social scoring, and real-time biometric identification systems.^{clxxvi}

High risk AI systems, on the hand, are divided into two categories: safety component of a product falling under the EU health and safety legislation (toy, aviation, cars, etc.), or systems deployed in specific areas such as biometrics identification, critical infrastructure, education, and employment. Systems in the category are required to register in an EU-wide database before the system is sold and put into service. Providers of such AI systems outside of the EU would have to conduct conformity assessments to ensure compliance.^{clxxvii} Other than that, High risk systems need to fulfill a range of requirements in various areas, including testing, technical, data training, amongst others.^{clxxix} Further, developers are required to set up reporting mechanisms for reporting serious incident to relevant authorities, such as those that may lead to damages to person’s health, death, or property.^{clxxx}

Limited risk systems, then, are identified as systems with limited risks of manipulation. Examples include deepfakes and chatbots. These systems would be subject to transparency obligations.^{clxxxii}

Finally, for AI systems with Minimal or Low risks, there are currently no obligations to fulfill any requirements. However, a code of conduct may be devised to encourage voluntary compliance of requirements for High risk AI systems.^{clxxxiii}

As of December 2023, the European Parliament has reached a provisional political agreement on the EU AI Act.^{clxxxiii} While the provisional text is not yet available, key provisions include placing obligations on “General Purpose AI models,” previously referred to as foundation models, and additional requirements for a subcategory with systemic risk. Notably, providers of General Purpose AI models may not face direct requirements but are asked to share technical information for high risk use cases. The EU AI Act is expected to enter into force early to mid-next year, with implantation periods varying by provision. Considered the first comprehensive legal framework to be introduced globally, the EU AI Act may lead countries around the world to adopt or pass formal legislation that aligns with the Act.^{clxxxiv} Similar observations can be made with the GDPR, which is widely adopted or referenced in data protection legislations in different jurisdictions.

Appendix C

Mapping of global AI governance initiatives against Korea's Digital Bill of Rights



Korea's approach to AI governance is congruent with global thinking and the technology industry at large.

Korea's approach to AI governance, as manifest in the five principles for a digital society of mutual prosperity in the Digital Bill of Rights, shares major commonalities with global initiatives delineated above.

Some noteworthy points of comparison include:

Principles as foundation The G7 Guiding Principles, OECD AI Principles, US Voluntary Commitments, and the Digital Bill of Rights all establish a foundation based on guiding principles that serve as a set of fundamental values to inform the responsible development and use of AI.

Risk-based approach as a key guiding principle The G7, OECD, and Korea's Digital Bill of Rights, as well as the US Executive Order, emphasize a risk-based approach to AI governance. They all acknowledge the dynamic nature of AI technologies and highlight the importance of monitoring for and mitigating risks throughout the lifecycle of AI systems from pre-development to post-deployment.

Ethical considerations All the global initiatives prioritize ethical considerations in AI governance in unison. Core values that are repeatedly emphasized include fairness, accountability, transparency, and human-centric approaches. Various follow-up measures to assess and implement these values are in the works. One of the most important objectives for all AI governance initiatives is to ensure the safety, security, and trustworthiness of AI systems. Especially the US, EU and Korea recognize that safety and trust are imperative from the perspective of industry promotion.

Multistakeholder engagement and partnership All the leading AI players in the global stage call for multistakeholder collaboration with industry partners to develop and refine the scope and methods of governing and managing AI systems. For instance, the UK at the AI Safety Summit sought leading AI stakeholders and parties to share their AI safety policies to promote transparency and an awareness of best practices. The G7 calls for prioritizing research and investment in risk-mitigating measures, while the US also directly commits to expand multistakeholder engagements for a holistic framework of AI governance.

International collaboration International cooperation is embedded in various principles, such as the G7, OECD, US Executive Order, and the Bletchley Declaration, as well as the Digital Bill of Rights. They recognize the global nature of AI-driven challenges, such as enhanced disparity from gaps in digital literacy, sustainability, and climate change, and call for international collaboration on risk identification and mitigation.

Table 1: Crosswalk of recent global AI policy developments against Korea’s Digital Bill of Rights

	Principle 1 Digital Freedom and Rights (Chapter 2)	Principle 2 Fair access and opportunities in the digital realm (Chapter 3)	Principle 3 Creation of a safe and trustworthy digital society (Chapter 4)	Principle 4 Encouragement of autonomous and creative digital innovation (Chapter 5)	Principle 5 Promotion of the well-being for all humankind (Chapter 6)
G7 AI Principles	Manage data quality and protect personal data and intellectual property (11)	Address global challenges (9)	Identify, evaluate, and mitigate risks across AI lifecycle (1) Monitor vulnerabilities, risks, and misuse after deployment and market placement (2) Responsible information sharing and reporting of incidents across organizations (4) Develop, implement, and disclose AI governance and risk management policies (5) Robust security controls (6) Reliable content authentication and provenance mechanisms (7) Manage data quality and protect personal data and intellectual property (11)	Responsible information sharing and reporting of incidents across organizations (4) Prioritize research to advance AI safety, security, and trust (8) Publicly report AI system’s capabilities and domains of use for transparency and accountability (3)	Address global challenges (9) Develop and adopt international technical standards (10)
US White House Executive Order	Advancing Equity and Civil Rights 1-3	Standing Up for Consumers, Patients, and Students 2: Shape AI’s potential to transform education Supporting Workers 1: Develop principles and best practices to mitigate the harms and maximize the benefits of AI for workers by addressing job displacement; labor standards; workplace equity, health, and safety; and data collection Promoting Innovation and Competition 1-3	New Standards for AI Safety and Security 1: Require that developers of the most powerful AI systems share their safety test results and other critical information with the US government New Standards for AI Safety and Security 2: Develop standards, tools, and tests to help ensure that AI systems are safe, secure, and trustworthy New Standards for AI Safety and Security 4: Protect Americans from AI-enabled fraud and deception by establishing standards and best practices for detecting AI-generated content and authenticating official content New Standards for AI Safety and Security 6: Order the development of a National Security Memorandum on safe, ethical, and effective use of AI in security Protecting Americans’ Privacy 1-4	Protecting Americans’ Privacy 1: Protect Americans’ privacy by prioritizing federal support for accelerating the development and use of privacy-preserving techniques Protecting Americans’ Privacy 2: Strengthen privacy-preserving research and technologies Standing Up for Consumers, Patients, and Students 1: Advance the responsible use of AI in healthcare and the development of affordable and life-saving drugs Promoting Innovation and Competition 1-3	Supporting Workers 1-2 Advancing American Leadership Abroad 1-3

The shared commitment to principles-based governance underscores the recognition that AI systems should align with a set of core values and ethical considerations. This approach allows for adaptability across diverse contexts while providing a common foundation for the responsible development and use of AI technologies globally. In the same vein, Korea’s Digital Bill of Rights aligns with several global initiatives. It also expands its scope by addressing a broader range of digital issues beyond AI, including digital literacy, disparities, and international cooperation.

US NIST AI Risk Management Framework

**Principle 1
Digital Freedom and Rights
(Chapter 2)**

GOVERN 3: Workforce diversity, equity, inclusion, and accessibility processes are prioritized in the mapping, measuring, and managing of AI risks throughout the lifecycle

**Principle 2
Fair access and opportunities in the digital realm (Chapter 3)**

AI presents significant risks, and efforts should be made to address issues such as human rights protection, transparency, fairness, accountability, safety, ethical considerations, and privacy.

**Principle 3
Creation of a safe and trustworthy digital society (Chapter 4)**

GOVERN 1: Policies, processes, procedures, and practices across the organization related to the mapping, measuring, and managing of AI risks are in place, transparent, and implemented effectively

GOVERN 2: Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible, and trained for mapping, measuring, and managing AI risks

GOVERN 4: Organizational teams are committed to a culture that considers and communicates AI risk

GOVERN 6: Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues

MEASURE 2: AI systems are evaluated for trustworthy characteristics

MANAGE 4: Risk treatments, including response and recovery, and communication plans for the identified and measured AI risks are documented and monitored regularly

AI should be designed, developed, deployed, and used in a manner that is safe, human-centric, trustworthy, and responsible.

AI presents significant risks, and efforts should be made to address issues such as human rights protection, transparency, fairness, accountability, safety, ethical considerations, and privacy.

Developers of powerful and potentially harmful AI systems have a strong responsibility for safety, including safety testing, evaluations, transparency, and accountability.

Special attention is required for highly capable general-purpose AI models and narrow AI that may cause harm, with a focus on understanding and mitigating risks, especially in cybersecurity, biotechnology, and disinformation.

**Principle 4
Encouragement of autonomous and creative digital innovation (Chapter 5)**

GOVERN 5: Processes are in place for robust engagement with relevant AI actors

Focus on identifying shared AI safety risks, building scientific understanding, developing risk-based policies, increasing transparency, and supporting an internationally inclusive network of scientific research on frontier AI safety.

All actors, including nations, international fora, companies, civil society, and academia, play a role in ensuring AI safety.

**Principle 5
Promotion of the well-being for all humankind (Chapter 6)**

MAP 5: Impacts to individuals, groups, communities, organizations, and society are characterized

AI's transformative opportunities should be harnessed for the greater good, including public services, human rights, environmental sustainability, and the achievement of UN Sustainable Development Goals.

Sustain an inclusive global dialogue, engage existing international fora, contribute to broader discussions, and continue research on frontier AI safety to harness the benefits of AI responsibly.

Focus on identifying shared AI safety risks, building scientific understanding, developing risk-based policies, increasing transparency, and supporting an internationally inclusive network of scientific research on frontier AI safety.

Appendix D

Best practices from Microsoft to build safe and trustworthy AI systems for a digital society of mutual prosperity

Responsible by Design: Outlining approaches to building AI systems for a digital society of mutual prosperity

For the above principle-based approach to AI governance to be effective, industry players need to embody and hold themselves accountable to such principles. Some leading AI companies, such as Microsoft, have been working extensively with different stakeholders to operationalize AI safety, and demonstrate accountability and responsibility.

Five-point blueprint for the public governance of AI

On May 25, 2023, Microsoft unveiled 'Governing AI: A Blueprint for the Future', presenting a comprehensive five-point blueprint for public AI governance.

A five-point blueprint for governing AI

- 01 **Implement and build** upon new government-led AI safety frameworks
- 02 **Require effective safety brakes** for AI systems that control critical infrastructure
- 03 **Develop a broader legal and regulatory framework** based on the technology architecture for AI
- 04 **Promote transparency** and ensure academic and public access to AI
- 05 **Pursue new public-private partnerships** to use AI as an effective tool to address the inevitable societal challenges that come with new technology

First, implement and build upon new government-led AI safety frameworks. Adopting and enhancing the AI Risk Management Framework developed by the U.S. National Institute of Standards and Technology (NIST) is important. Governments can boost alignment with global standards by designing their respective AI governance frameworks in alignment with the NIST framework. Microsoft's voluntary commitments at the White House (more info below) is an extension of this effort.

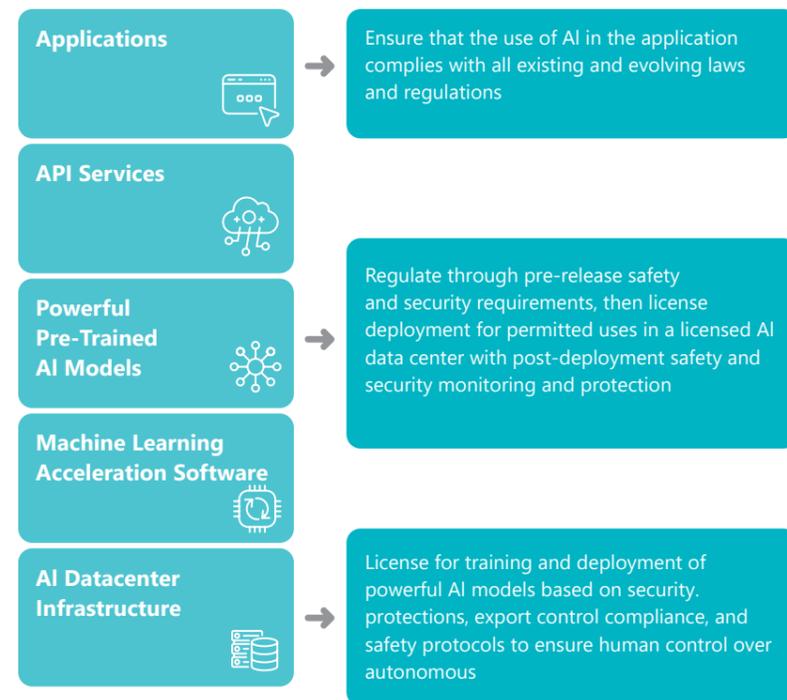
Second, require effective safety-brakes for AI systems that control critical infrastructure. The blueprint advocates for mandatory safety mechanisms in AI systems overseeing critical infrastructure. These mechanisms are akin to safety brakes in various technologies, aiming to maintain human oversight and ensure system resilience. High risk AI systems must be defined, fail-safe designs mandated, and such systems should be operated in licensed AI data centers for additional security.

Third, develop a broad legal and regulatory framework based on the technology architecture for AI. Developing a legal and regulatory framework that reflects AI's technological architecture is vital. This involves assigning regulatory responsibilities across different layers of the AI technology stack and applying a "Know Your 3C's" principle (Cloud, Customer, Content) to manage risks associated with AI deployment and usage. From Applications to Infrastructure, there needs to be a regulatory architecture tailored and appropriate for each.

The technology stack for AI foundation models

Applications 	Software programs where the output of an AI model is put to work
API Services 	APIs (Application Program Interfaces), or endpoints, through which applications access pre-trained models
Powerful Pre-Trained AI Models 	Pre-trained models like GPT-4 that can be used to solve similar problems without starting from scratch
Machine Learning Acceleration Software 	Software that speeds up the process of developing and deploying large AI models
AI Datacenter Infrastructure 	Advanced supercomputing infrastructure, including clusters of advanced GPUs (Graphics Processing Units) with high bandwidth network connections

A proposed AI regulatory architecture



Fourth, promote transparency and ensure academic and nonprofit access to AI. Advancing transparency is crucial for broadening access to AI resources. Microsoft committed to an annual AI transparency report and other steps to expand transparency for our AI services. It is also critical to expand access to AI resources for academic research and the nonprofit community. Unless academic researchers can obtain access to substantially more computing resources, there is a real risk that scientific and technological inquiry will suffer, including relating to AI itself.

Fifth, pursue new public-private partnerships to use AI as an effective tool to address the inevitable societal challenges that come with new technology. Key to harnessing AI's benefits and mitigating its risks is robust collaboration between the public and private sectors. Such partnerships can protect democracy, promote inclusive growth, and contribute to sustainability.

Voluntary commitments from leading AI companies to manage potential risks of AI

Prior to the issuance of the White House Executive Order, seven leading AI companies (including Microsoft, Google, Amazon, Meta, and OpenAI) committed to a set of Voluntary Commitments with the US government with the objective of demonstrating safe, security and trustworthy development of AI technology.^{clxxxv} This commitment is anchored in three key principles: safety, security, and trust.

In brief, safety refers to committing to internal and external security testing by experts prior to public use. This serves to guard against AI risks such as bio, cyber and national security risks. The commitment also encourages the

sharing of best practices and risk management attempts across a range of different actors, which include government, academia, and civil society.

In terms of security, the commitment focuses on investing in cybersecurity and insider threat safeguards with consideration to security risks. It further focuses on robust reporting and discovery by third parties to facilitate the identification and management of vulnerabilities in AI systems.

Finally, in terms of earning public trust, the key commitment under this principle is the development of technical mechanisms, such as watermarking systems, to inform the public of AI-generated content. The commitment also calls for companies to publicly discuss AI systems' capabilities and limitations; as well as invest in research to avoid AI harm and commit to more advanced AI systems to tackle global challenges.

The securing of this first wave for voluntary commitment from AI companies had attracted additional companies to sign on.^{clxxxvi} The expressed support from the private sector had laid a sound foundation for the US government to push forward with the White House Executive Order.

Microsoft Commits to Advancing Safe, Secure, and Trustworthy AI

To expand its safe and responsible AI practices, Microsoft endorsed the new voluntary commitments initiated by the Biden-Harris administration, aimed at ensuring the safety, security, and trustworthiness of advanced AI systems. These voluntary commitments are guided by principles of safety, security, and trust and address risks associated with advanced AI models and are aligned with existing US government work, such as the NIST AI Risk Management Framework and the Blueprint for an AI Bill of Rights.

Going a step further, Microsoft has committed to strengthening the ecosystem by backing a National AI Research Resource pilot and advocating for a national registry of high risk AI systems to enhance transparency and accountability. To develop more reliable AI systems benefiting customers and society at large, Microsoft is supporting the widespread implementation of the NIST AI Risk Management Framework, and the adoption of cybersecurity practices tailored to the unique risks of AI.



Alignment of Our Efforts with the White House Voluntary AI Commitments

Safe

White House Voluntary Commitments:

Companies choose to conduct red-teaming, share trust and safety information, and help people identify AI-generated content

Microsoft Commitments:

- Test our systems using red-teaming and systematic measurements
- Contribute to industry efforts to develop evaluation standards for emerging safety and security issues
- Implement provenance tools to help people identify AI-generated audio or visual content
- **Implement the NIST AI Risk Management Framework**
- **Implement robust reliability and safety practices for high-risk models & applications**

Secure

White House Voluntary Commitments:

Companies choose to make investments to protect unreleased model weights, and incent the responsible disclosure of AI system vulnerabilities

Microsoft Commitments:

- Ensure that the cybersecurity risks of our AI products and services are identified and mitigated
- Participate in an approved multistakeholder exchange of threat information
- **Support the development of a licensing regime for highly capable models**
- **Support the development of an expanded know-your-customer' concept for AI services**

Trustworthy

White House Voluntary Commitments:

Companies choose to be transparent about system capabilities and limitations, prioritize research on societal risks, and develop and deploy AI systems for the public good

Microsoft Commitments:

- Release an annual transparency report on the governance of our responsible AI program
- Design our AI systems so that people know when they are interacting with an AI system and be transparent about system capabilities and limitations
- Increase investment in our academic research programs
- **Collaborate with the National Science Foundation to explore a pilot project to stand up the National AI Research Resource**
- **Support the development of a national registry of high-risk AI systems**

*red denotes our additional commitments

AI Safety Summit

Apart from the Bletchley Declaration, the AI Safety Summit also saw AI companies, including Microsoft, working with governments to reach an agreement on testing AI models.^{clxxxvii} Under the voluntary agreement, leading developers agree to undertake responsibilities to “create and execute safety testing, employing evaluations, transparency, and other suitable measures”^{clxxxviii} to mitigate risks and manage vulnerabilities. Regulators and developers, as per the agreement, would be enabled to work together on testing the safety of new AI models pre- and post-deployment.

To operationalize the agreement, developers, and participating governments (including the US, UK, EU, etc.) would be able to undertake testing led by the AI Safety Institute.

Leading tech companies should lead by example by implementing and operationalizing responsible AI in the design of their AI systems.

Leading tech companies play a pivotal role in shaping the responsible landscape of AI. To truly lead by example, they must take proactive steps to implement and operationalize responsible AI principles in the very fabric of their AI systems. This section unpacks the approaches taken by Microsoft such as establishing responsible guidelines for AI design, robust governance structures, and transparent accountability mechanisms. These companies can inspire the broader tech industry to adopt responsible AI practices and help set a standard that benefits society.

Microsoft's Responsible AI Journey

In 2016, Microsoft embarked on its responsible AI journey under Satya Nadella's vision. Taking lessons from long-standing, cross-company commitments to privacy, security, and accessibility, they realized that responsible AI must be supported by the highest levels of leadership in the company and championed at every level across Microsoft. Core AI principles were established, leading to the formation of the Aether Committee in 2018. This committee, consisting of experts, contributed to the creation and adoption of AI principles. Microsoft furthered its commitment in 2019 with the establishment of the Office of Responsible AI, overseeing governance and crafting the Responsible AI Standard for actionable guidance in building AI systems. At the working level, core teams within engineering, research, and policy play critical roles to advance responsible AI across the company, each bringing a set of unique



skills. Responsible AI roles are also embedded within product, engineering, and sales teams by the appointment of Responsible AI Champions by leadership. At the next level, the Responsible AI Council is a forum for leadership alignment and accountability in implementing Microsoft's responsible AI program. At the highest level, the Environmental, Social, and Public Policy Committee of the Microsoft Board provides oversight of their responsible AI program.

The Anatomy of the Responsible AI Standard

Principles		Which enduring values guide our responsible AI work?
Goals		What are the outcomes that we need to secure?
Requirements		What are the steps we must take to secure the goals?
Tools and practices		Which aids can help us meet the requirements?



Case study: How Microsoft applies their responsible AI approach to the new Bing

In February 2023, Microsoft launched the new Bing, an AI-enhanced web search experience. It supports users by summarizing web search results and providing a chat experience. Users can also generate creative content, such as poems, jokes, letters, and, with Bing Image Creator, images. The new AI-enhanced Bing runs on a variety of advanced technologies from Microsoft and OpenAI, including GPT-4, a cutting-edge large language model (LLM) from OpenAI. Responsible AI teams across Microsoft worked with GPT-4 for months prior to its public release by OpenAI to develop a customized set of capabilities and techniques to join this cutting-edge AI technology and web search in the new Bing.

In preparing for the launch, Microsoft harnessed the full power of their responsible AI ecosystem. The new Bing experience has been developed in line with Microsoft's AI Principles, Microsoft's Responsible AI Standard, and in partnership with responsible AI experts across the company,



including Microsoft's Office of Responsible AI, their engineering teams, Microsoft Research, and the Aether Committee.

Guided by their AI Principles and Responsible AI Standard, they sought to identify, measure, and mitigate potential harms and misuse of the new Bing while securing the transformative and beneficial uses that the new experience provides. The next section describes their approach, highlighting best practices companies can follow to implement a responsible AI approach.

Identify

At the model level, Microsoft's work began with exploratory analyses of GPT-4 in the late summer of 2022. This included conducting extensive red teaming in collaboration with OpenAI. This testing was designed to assess how the latest technology would work without any additional safeguards applied to it. The specific intention was to produce harmful responses (responses are outputs from the AI system—in this case, a large language model—and may also be referred to as "completions," "generations," and "answers"), to surface potential avenues for misuse, and to identify capabilities and limitations. The combined learnings advanced OpenAI's model development, informed Microsoft's understanding of risks, and contributed to early mitigation strategies for the new Bing.

In addition to model-level red teaming, a multidisciplinary team of experts conducted numerous rounds of application level red teaming on the new Bing AI experiences before making them available in their limited release preview. This process helped them better understand how the system could be exploited by adversarial actors and improve their mitigations. Non-adversarial testers also extensively evaluated new Bing features for shortcomings and vulnerabilities.

Measure

Red teaming can surface instances of specific harms, but in production, users will have millions of different kinds of conversations with the new Bing. Moreover, conversations are multi-turned and contextual, and identifying harmful responses within a conversation is a complex task. To better understand and address the potential for harms in the new Bing AI experiences, they developed additional responsible AI metrics specific to those new AI experiences for measuring potential harms like jailbreaks, harmful content, and ungrounded content. They also enabled measurement at scale through partially automated measurement pipelines.

Their measurement pipelines enable them to rapidly perform measurement for potential harms at scale, testing each change before putting it into production. As they identify new issues through the preview period and beyond, as well as ongoing red teaming, they continue to expand the measurement sets to assess additional harms.

Mitigate

As they identified and measured potential harm and misuse, they developed additional mitigations to those used for traditional search. Some of those include:

- ▶ **Preview period, phased release.** Their incremental release strategy has been a core part of how they move their technology safely from the labs into the world, and they are committed to a deliberate, thoughtful process to secure the benefits of the new Bing. Limiting the number of people with access during the preview period allowed us to discover how people use the new Bing, including how people may misuse it, before broader release. They continue to make changes to the new Bing daily to improve product performance, improve existing mitigations, and implement new mitigations in response to our learnings.
- ▶ **AI-based classifiers and metaprompting to mitigate harm or misuse.** The use of LLMs may produce problematic content that could lead to harm or misuse. Classifiers and metaprompting are two examples of mitigations that have been implemented in the new Bing to help reduce the risk of these types of content. Classifiers classify text to flag different types of potentially harmful content in search queries, chat prompts, or generated responses. Flags lead to potential mitigations, such as not returning generated content to the user, diverting the user to a different topic, or redirecting the user to traditional search. Metaprompting involves giving instructions to the model to guide its behavior. For example, the metaprompt may include a line such as “communicate in the user’s language of choice.”
- ▶ **Grounding in search results.** The new Bing is designed to provide responses supported by the information in web search results when users are seeking information. For example, the system is provided with text from the top search results and instructions via the metaprompt to ground its response. However, in summarizing content from the web, the new Bing may include information in its response that is not present in its input sources. In other words, it may produce ungrounded results. They have taken several measures to mitigate the risk that users may over-rely on ungrounded generated content in summarization scenarios and chat experiences. For example, responses in the new Bing that are based on search results include references to the source websites for users to verify the response and learn more. Users are also provided with explicit notice that they are interacting with an AI system and are advised to check the web result source materials to help them use their best judgment.
- ▶ **Limiting conversational drift.** During the preview period, they learned that very long chat sessions can result in responses that are repetitive, unhelpful, or inconsistent with new Bing’s intended tone. To address this conversational drift, they limited the number of turns (exchanges which contain both a user question and a reply from Bing) per chat session,



Case Study: How Microsoft empowers customers on their responsible AI journey - Azure’s online safety measures

until they could update the system to better mitigate the issue.

- ▶ **AI disclosure.** The new Bing provides several touchpoints for meaningful AI disclosure, where users are notified that they are interacting with an AI system as well as opportunities to learn more about the new Bing.

Microsoft’s approach to identifying, measuring, and mitigating harms will continue to evolve as they learn more—and as they make improvements based on feedback gathered during the preview period and beyond.

One of Microsoft’s most important responsible AI commitments is to help customers on their responsible AI journey by sharing their learnings with them. They understand that their efforts alone are not enough to secure the societal gains envisioned when responsible AI practices are adopted.

As part of this commitment, Microsoft provides transparency documentation for their platform AI services in the form of Transparency Notes to empower their customers to deploy their systems responsibly. Transparency Notes communicate in clear, everyday language the purposes, capabilities, and limitations of AI systems so that their customers can understand when and how to deploy their platform technologies. They also identify use cases that fall outside the solution’s capabilities and the Responsible AI Standard. Transparency Notes fill the gap between marketing and technical documentation, proactively communicating information that their customers need to know to deploy AI responsibly.^{clxxxix}

Customers also need practical tools to operationalize responsible AI practices. Over the years, responsible AI research at Microsoft has led to the incubation of tools such as Fairlearn and InterpretML. The collection of tools has grown in capability, spanning many facets of responsible AI practice including the ability to identify, diagnose, and mitigate potential errors and limitations of AI systems. Since their original conception within Microsoft, these tools continue to improve and evolve externally through the contributions of active open-source communities. The collection of tools can be found under the Responsible AI Toolbox GitHub repository. Their latest tool, which is in preview, is Azure Content Safety which helps businesses create safer online environments and communities through models that are designed to detect hate, violent, sexual, and self-harm content across languages in both images and text.



Building on the Responsible AI Toolbox, Microsoft's responsible AI program has invested in integrating some of the more mature responsible AI tools directly into Azure Machine Learning so that their customers will also benefit from the development of engineering systems and tools. The collection of capabilities, known as the Responsible AI Dashboard, offers a single pane of glass for machine learning practitioners and business stakeholders to debug models and make informed, responsible decisions as they build AI systems or customize existing ones. Some of their latest features added in preview include support for text and image data that enables users to evaluate large models built with unstructured data during the model-building, training, and evaluation stages, and Prompt Flow, which provides a streamlined experience for prompting, evaluating, and tuning large language models, including on measurements such as groundedness.

Responsible AI built into Azure Machine Learning

Fairness

Assess fairness and mitigate fairness issues to build models for everyone.



Explainability

Understand model predictions by generating feature importance values for your model.



Causal analysis

Estimate the effect of a feature on real-world outcomes.



Error analysis

Identify dataset cohorts with high error rates and visualize error distribution in your model.



Counterfactuals

Observe feature perturbations and find the closest datapoints with different model predictions.



Responsible AI scorecard

Get a PDF summary of your Responsible AI insights to share with your technical and nontechnical stakeholders to aid in compliance reviews.



Prompt Flow

Create workflows for large language-based applications to simplify prompt building, evaluation, and tuning.



Azure Content Safety

Detect hate, violent, sexual, and self-harm content across languages in both images and text.



The community involved in developing, evaluating, and using AI expands beyond their direct customers. To serve this broad ecosystem, they publicly share key artifacts from their responsible AI program, including their Responsible AI Standard, Impact Assessment template, and collections of cutting-edge research. Their digital learning paths further empower leaders to craft an effective AI strategy, foster an AI-ready culture, innovate responsibly, and more.^{cxv}

Protecting customers while respecting copyrights

In September 2023, Microsoft introduced the Copilot Copyright Commitment to address concerns regarding copyright issues associated with their AI-powered Copilots.^{cxvi} This commitment assures customers that they can use Copilot services and the generated content without worrying about potential copyright claims. Microsoft takes responsibility for any legal risks arising from copyright infringement challenges on the customer's end.

Building upon the AI Customer Commitments, Microsoft introduced updates to the Copilot Copyright Commitment in December 2023, unveiling the Customer Copyright Commitment (CCC). This commitment underscores Microsoft's dedication to shielding customers from specific third-party intellectual property claims linked to content generated by paid versions of Microsoft commercial Copilot services (including Windows Copilot when signed in with a work ID), Microsoft Copilot (formerly, Bing Chat Enterprise), the Azure OpenAI Service, and other generative AI services with configurable Metaprompts and other safety systems.^{cxvii}

Microsoft's Customer Copyright Commitment¹

Microsoft's obligation to defend Customer against third-party IP claims under Customer's volume licensing agreement will apply to Customer's use or distribution of Output Content of a Covered Product if all the following additional conditions are met:

- 01 While using the Covered Product to produce the Output Content that is the subject of the claim, Customer must not have disabled, evaded, disrupted, or interfered with the content filters, restrictions in Metaprompts, or other safety systems that are part of the Covered Product.
- 02 Customer does not modify, use, or distribute the Output Content in a manner that it knows, or should know, is likely to infringe or misappropriate any proprietary right of a third party.
- 03 Customer has sufficient rights to use the Input in connection with the Covered Product, including, without limitation, any Customer Data that Customer used to Customize the model that produced the Output Content that is the subject of the claim.
- 04 The claim does not allege that the Output Content, as used in commerce or the course of trade, violates a third party's trademark or related rights.
- 05 For Azure OpenAI Service and any Microsoft Generative AI Service with configurable Metaprompts or other safety systems, Customer also must have implemented all mitigations required by the Azure OpenAI Service documentation in the offering that delivered the Output Content that is the subject of the claim.

¹ Microsoft (2023). Customer Copyright Commitment Required Mitigations. Available at: <https://learn.microsoft.com/en-us/legal/cognitive-services/openai/customer-copyright-commitment>

The program is rooted in Microsoft's commitment to stand behind its customers, ensuring that using their products doesn't lead to legal complications. It also acknowledges the concerns of authors and aims to strike a balance between advancing AI technology for societal benefits and respecting copyright. Microsoft has implemented filters and technologies to reduce the likelihood of its products producing infringing content, and customers are incentivized to use these technologies to address copyright concerns. The Customer Copyright Commitment applies to paid versions of Microsoft's commercial Copilot services (including Windows Copilot when signed in with a work ID), Microsoft Copilot (formerly Bing Chat Enterprise) and the Azure OpenAI Service.

Appendix E

Case studies on the potential impact of the AI industry on the Korean economy

In Korea, a nation at the forefront of technological innovation, the emergent landscape of AI is not merely a testament to human ingenuity but a vital tool in addressing some of its most pressing societal challenges.

As this section will demonstrate, AI offers transformative solutions across a spectrum of issues in Korean society. This section will go issue by issue and examine ways in which Korean companies have leveraged the power of AI to respond to real world issues. These issues include a focus on:

- 01 AI in Healthcare
- 02 AI in Education
- 03 AI for Economic Development
- 04 AI in Public Services

AI's implementation across various sectors marks a significant leap forward, impacting everything from healthcare and education to social discourse and economic development. AI's smart applications are enhancing patient care, personalizing education, moderating online communities, and driving business innovation. In public services, AI is refining emergency response and urban planning, demonstrating the nation's dedication to embracing AI for comprehensive progress.

The following sections will synthesize insights from diverse case studies within the Korean context, mapping out the strategic deployment of AI across these identified areas:

AI for Healthcare

AI-Driven Healthcare Innovation in Korea's Aging Society

Korea faces an aging population challenge, with the elderly expected to constitute 24% of the populace by 2030, elevating the prevalence of age-related diseases. To address this, Professor Sun Choi's team at Ewha Womans University, in collaboration with Microsoft Research's AI4Science in Asia team, have embarked on an innovative AI-driven research initiative. Their focus is on the development of TamGent, a generative model for molecule generation targeting G Protein-Coupled Receptors (GPCRs), which are critical in age-related disorders. Utilizing the transformative capabilities of AI in drug discovery, TamGent can generate novel compounds by directly targeting protein complexes, thus bypassing conventional compound library searches. This approach significantly accelerates the drug development process, potentially reducing costs and enhancing medical treatments for the elderly. This research is pivotal in addressing the national financial burden of healthcare for an aging population, offering a promising solution for enhancing the health and wellbeing of the elderly in South Korea.

Advancing Drug Development and Clinical Trials in Korea Through AI

In a project spearheaded by Professor Howard Lee of Seoul National University College of Medicine, and supported by the Microsoft Research Lab, work has been undertaken to confront two prevalent problems in drug development, particularly impactful in the context of South Korea's



aging population. The first issue is the inefficiency in drug repurposing, a process crucial for discovering new uses for existing drugs. The research team employs AI to navigate the complexities of drug indication and drug characteristics data, aiming to mitigate shortcut learning and improve the accuracy of drug-disease associations. This approach promises to streamline the drug repurposing process, offering a more efficient pathway to identifying effective treatments.

The second part of the study addresses the limitations of current Eligibility Criteria (EC) in clinical trials. Traditional EC often limits the diversity and representativeness of trial participants, which is particularly problematic for trials targeting elderly patients. To tackle this, the research introduces an AI-driven method, CReSE, which utilizes contrastive learning and text rephrasing with Large Language Models (LLMs) to generate more inclusive and clinically relevant EC. This innovation is expected to enhance the inclusivity and safety of clinical trials, ensuring they are more representative of the diverse patient populations, including the elderly. These advancements in AI application in drug repurposing and clinical trial design are pivotal in improving healthcare outcomes, particularly for the elderly in South Korea's aging society.

AI in Education

MS AI School^{exdiii} : Preparing a future-ready workforce for Korea's tomorrow

AI is revolutionizing the educational landscape, particularly in the domain of professional training and job readiness. MS AI School, an initiative by Microsoft Philanthropies, exemplifies this transformation by providing AI job training to bridge the employment gap in Korea, where a significant shortage of skilled AI professionals persists.

The program localizes the Microsoft AI School model, initiated in France, to Korean needs, offering comprehensive training courses with incentives and allowances, facilitated by partnerships with government and industry bodies. This initiative not only equips job seekers with AI skills but also directly addresses the workforce deficit highlighted by the Software Policy and Research Institute (SPRI).

Personal testimonies illustrate the profound impact of AI on education. For instance, Lee Seung-Yoon transitioned from a background in cultural heritage conservation to AI and big data after participating in the MS AI School. The intensive six-month training program provided him with a broad understanding of AI, from basic programming to advanced concepts like deep learning. His success story underscores the school's capacity to empower individuals from diverse academic backgrounds to pivot into the burgeoning field of AI.

The quality of instruction at MS AI School is frequently lauded by its alumni. Competent instructors with real-world experience offer high-quality lectures, which are considered a blessing in an era that increasingly favors solitary study.

Participants like Hwang Hye-jung value the school's direct association with Microsoft and the opportunity to immerse themselves in AI, citing the use of high-specification Microsoft Cloud Azure VMs and the chance to gain Microsoft-certified Azure certifications as key benefits. The program's robust curriculum receives high marks from industry insiders, indicating a close alignment with both theoretical knowledge and practical application.

MS AI School fosters a collaborative learning environment through team projects, which cultivates 'team play' and offers various synergistic benefits, such as networking and employment insights. This team-based approach expands the educational experience and enhances the joy of learning. Trainees like Ms. Hwang, aspiring to shine in the male-dominated IT industry, see the program as a steppingstone towards immediate employment in specialized fields like computer vision. Despite concerns about a contracting job market, participation in MS AI School is viewed as a valuable asset, likely to spark conversations with HR managers. This focus on AI education is timely, considering the World Economic Forum's prediction of job displacement due to technological innovation, underscoring the increasing importance of AI expertise in the job market.

AI for Economic Development

AI-Powered Revival of Korean Literary Heritage^{CXCIV}

Dr. Wayne de Fremery, a Korean studies scholar at Sogang University, is at the forefront of integrating AI into education and cultural heritage through his work with Azure and machine learning technologies. His project focuses on the digital resurrection of historical Korean texts that are dense with over 11,000 unique alphabetic characters and ancient Chinese ideograms, akin to the complexity of sixteenth-century English poetry.

The challenge for Korean students today is not just reading but comprehending these classical texts. Dr. de Fremery's AI endeavor aims to decode these texts, making them accessible and intelligible to the modern reader. The use of machine learning on Azure enables the parsing of intricate patterns within the literature, providing nuanced translations and facilitating a better understanding of Korea's literary history.

This technological application has substantial implications for educational engagement and cultural preservation. It connects young Koreans with their heritage and offers a model for using AI to unlock cultural treasures globally, revolutionizing the ways in which historical legacies are preserved and studied.

Bentley Systems with Doosan Heavy Industries^{CXCV} : AI-based infrastructure management solutions at a scale

Bentley Systems, a leader in infrastructure engineering software, has been utilizing the power of digital twins and AI on the Microsoft Azure platform to enhance global infrastructure management. These advanced technological solutions integrate disparate data sources, facilitating insight generation and informed decision-making for industry professionals.

With the use of digital twins, Bentley Systems has transformed the approach to infrastructure development and maintenance, providing a dynamic and data-enriched representation of physical systems. This has proven economically beneficial for many entities. Among others, Bentley Systems' collaboration with Doosan Heavy Industries and Construction in Korea exemplifies the application of AI in promoting economic development through digital twins. Doosan, a company that provides power generation alongside engineering and construction services in Korea, utilized Bentley Systems' technology, underpinned by Microsoft Azure, to create an advanced digital twin for its wind farms. This virtual model is a sophisticated blend of real-time and historical data, including Internet of Things (IoT) and weather data, integrated through Azure's powerful cloud computing capabilities.

The digital twin developed for Doosan's wind farms is a prime example of technological innovation aiding in economic development. It allows for remote monitoring of the turbines, delivering real-time operational efficiencies. The AI component employs physics-based and machine learning algorithms to predict energy production based on variable weather conditions. These predictions enable Doosan to fine-tune the performance of each turbine, maximizing energy output, which directly translates to increased economic benefits and operational savings.

The insights gained from the digital twin are not only instrumental in the real-time optimization of energy production but also feed into the design process of future wind turbines. By accurately predicting output and understanding the behavior of turbines under various conditions, Doosan can enhance the design and efficiency of its subsequent models. This continuous improvement cycle is vital for maintaining competitiveness in the energy sector and supports Korea's broader economic development goals by bolstering its renewable energy capabilities.

Bentley Systems, through its partnership with Microsoft, has provided Doosan Heavy Industries with a sophisticated digital solution that has tangible economic advantages. This initiative underscores the transformative impact AI-infused digital twins can have on industry practices, infrastructure management, and economic advancement, particularly in the realm of sustainable energy solutions in South Korea and beyond.

Lotte Hotel^{CXCVI} : Focus on the aspects of hospitality that matters through automation

Microsoft Korea has embarked on a venture with Lotte Hotel, aiding in the creation of a hyper-automated work environment leveraging the Power Platform. This innovative move has enabled the automation of tasks without requiring significant IT expertise, leading to the delegation of repetitive duties to AI-driven robots. This digital transformation across 17 Lotte Hotels has culminated in an annual time saving of approximately 10,000 hours, showcasing the practical benefits of automation in the hospitality industry.

The Power Platform is a suite of Microsoft's low-code tools that democratize the process of app development and business automation. The platform, which includes Power Apps, Power Automate, Power BI, and Power Pages, has allowed Lotte Hotel staff to efficiently automate necessary tasks. Notably, the automation of the reservation system, which integrates external booking information into the hotel's system, has not only saved time but also reduced manual entry errors significantly.

To encourage the use of Power Platform within its business operations, Lotte Hotel has fostered an environment of innovation among its employees. Through the establishment of an RPA council, it promotes the sharing of knowledge and experience in automation, running contests, and rewarding initiatives that result in cost savings. These efforts are part of Lotte Hotel's strategy to embed a culture of continuous improvement and productivity through technological empowerment.

Lotte Hotel's strategic expansion plan aims to extend the application of Power Platform to various business challenges, including customer service and operational management. By enabling employees to develop apps and web pages and providing access to advanced AI technologies such as chatbots and Azure OpenAI Service, Lotte Hotel is poised to experience a significant enhancement in work efficiency. Endorsements from Junwoo Kim, the Digital Strategy Director at Lotte Hotel, and Ohsung Kwon from Microsoft Korea, underscore the effectiveness of Power Platform in streamlining business processes, while also maintaining high standards of customer satisfaction and security.

AI for Public Services

AI Enhancements in Seoul's Emergency Services^{cxvii}

In an innovative approach to public safety, a university in Seoul has spearheaded a project utilizing AI to predict fire outbreaks with remarkable accuracy. Professor Jae Seung Lee and his students from Hongik University employed machine learning algorithms to analyze data from the Seoul Fire Department. Their model, developed using Azure Machine Learning Studio, achieves a 90% accuracy rate in forecasting fire incidents, a significant step in public service enhancement.

This breakthrough in predictive modelling has directly influenced the efficiency of fire response teams. By identifying areas with a higher likelihood of fire incidents, the fire service has been able to adjust patrol routes and allocate resources more effectively. The improved deployment of fire crews, especially in high-risk districts, has led to quicker response times, potentially saving lives and reducing property damage. Surprising insights from the data revealed newer, busier districts such as Gangnam to be more susceptible to fires, challenging prior assumptions and informing strategic planning.

The project has also been pivotal in terms of urban management and public trust. Recognizing the sensitivity of the data, Professor Lee

recommended the use of a secure Microsoft virtual machine to ensure citizens' privacy was upheld, a move that fostered trust between the university team and the Fire Department. This secure data handling has allowed the project to flourish and be received positively by city officials, keen on using cutting-edge technology to improve public welfare and resource management.

Looking forward, the success of the fire prediction model has set the stage for its application to other civic challenges, including crime prevention and traffic management. The ongoing project on wheelchair accessibility is just one example of the university's commitment to using AI for the betterment of the city. Professor Lee's focus on equipping students with the necessary skills and domain knowledge highlights the potential for AI to transform public services in Seoul, enhancing not just emergency responses but also the broader aspects of urban living.



References

- i MSIT (2019). National Strategy for Artificial Intelligence. Available at: <https://www.msit.go.kr/bbs/view.do?sCode=eng&nttSeqNo=9&bbsSeqNo=46&mId=10&mPid=9>
- ii MSIT (2022). “Korea to Come up with the Roadmap of Digital ROK, Realizing the New York Initiative.” Available at: <https://www.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=4&mPid=2&bbsSeqNo=42&nttSeqNo=742>
- iii Korea.net (2022). “Korea jumps 4 spots to 8th in digital competitiveness rankings.” Available at: <https://www.korea.net/NewsFocus/Sci-Tech/view?articleId=221834>
- iv MSIT (2022). “국가전략기술 임무중심 전략로드맵 수립.” Available at: <https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=238&pageIndex=15&bbsSeqNo=94&nttSeqNo=3183427&searchOpt=ALL&searchTxt=>
- v Korea IT News (2023). “국내 AI 시장 연 14.9% 성장…2027년 4.4조 전망.” Available at: <https://www.koit.co.kr/news/articleView.html?idxno=112323>
- vi Microsoft and KCCI (2023). The Economic Impact of Generative AI: The Future of Work in South Korea. Available at: https://www.korcham.net/nCham/Service/Economy/appl/KcciReportDetail.asp?SEQ_NO_C010=20120936516&CHAM_CD=B001
- vii MSIT (2022). “Korea to Come up with the Roadmap of Digital ROK, Realizing the New York Initiative.”
- viii TechM (2023). “영국 AI 안전성 정상회의 참석한 이종호 과기정통부 “적절한 AI 신뢰성·안전성 확보해야.” Available at: <https://www.techm.kr/news/articleView.html?idxno=116260>
- ix Salesforce (2023). Asia Pacific AI Readiness Index. Available at: https://www.salesforce.com/content/dam/web/en_sg/www/documents/pdf/salesforce_ai_readiness_index_2023.pdf
- x MSIT (2022). “Korea to Come up with the Roadmap of Digital ROK, Realizing the New York Initiative.”
- xi MSIT (2023). “디지털 권리장전으로 OECD의 디지털 규범 논의를 이끌어간다.” Available at: https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=238&bbsSeqNo=94&nttSeqNo=3183648&formMode=&pageIndex=2&searchCtgy=&searchCtgy2=&searchCtgy3=&RLS_YN=&searchOpt=ALL&searchTxt=
- xii MSIT (2023). “과기정통부, ‘인공지능 안전성 정상회의’ 참석 성과 발표.” Available at: <https://www.msit.go.kr/bbs/view.do>
- xiii KDI (KDI (2023). Digital Bill of Rights. Available at: <https://eiec.kdi.re.kr/policy/materialView.do?num=2431742023>. 디지털 권리장전. 출처: <https://eiec.kdi.re.kr/policy/materialView.do?num=243174>
- xiv Digital Public Sphere (2023). “Digital Bill of Rights.”
- xv KISDI (2022). 인공지능 윤리기준 실천을 위한 자율점검표.; TTA (2022). 신뢰할 수 있는 인공지능 개발안내서.; KISDI (2022). 인공지능 윤리교육 교재.
- xvi Policy Briefing (2023). “민간 자율 인공지능 신뢰성 검·인증 설명회 개최.” Available at: <https://www.korea.kr/briefing/pressReleaseView.do?newsId=156597130#pressRelease>
- xvii Korea National Assembly (2023). AI Responsibility and Regulation Bill. Available at: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_W2E3F0D8D0C1A0B913J1I21G4F0G1
- xviii PIPC (2023). “신뢰기반 인공지능 데이터 규범, 첫 발을 떼다.” Available at: <https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=9083>
- xix OECD AI Policy Observatory (2023). OECD AI Principles overview.
- xx Ministry of Foreign Affairs, Japan (2023). Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System. Available at: <https://www.mofa.go.jp/files/100573471.pdf>
- xxi The White House (2023). “Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.”
- xxii NIST (2023). AI Risk Management Framework.
- xxiii Cloud Security Alliance (2023). NIST’s AI Risk Management Framework Explained. Available at: <https://cloudsecurityalliance.org/blog/2023/08/30/nist-s-ai-risk-management-framework-explained/>
- xxiv GOV.UK (2023). Chair’s Summary of the AI Safety Summit 2023, Bletchley Park.

References

- xxv Ibid.
- xxvi Microsoft (2023). Ongoing Russia cyberattacks targeting Ukraine. Available at: <https://blogs.microsoft.com/on-the-issues/2023/06/14/russian-cyberattacks-ukraine-cadet-blizzard/>
- xxvii The White House (2023). “Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI.” Available at: <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>
- xxviii Reuters (2023). “Adobe, others join voluntary US scheme to manage AI risks.” Available at: <https://www.reuters.com/sustainability/adobe-others-join-white-houses-voluntary-commitments-ai-2023-09-12/>
- xxix Microsoft (2018). Getting ready for General Data Protection Regulation (GDPR). Available at: <https://www.microsoft.com/en-us/industry/microsoft-in-business/business-transformation/2018/03/29/getting-ready-for-general-data-protection-regulation-gdpr/>; Microsoft (2022). Microsoft’s framework for building AI systems responsibly. Available at: <https://blogs.microsoft.com/on-the-issues/2022/06/21/microsofts-framework-for-building-ai-systems-responsibly/>
- xxx Digital Public Sphere (2023). “Digital Bill of Rights.”
- xxxi Law Times (2020). “‘인공지능 (AI) 법제정비단’ 민간위원 위촉.” Available at: <https://www.lawtimes.co.kr/news/161045?serial=161045>
- xxxi MSIT (2020). “인공지능 시대를 준비하는 법·제도·규제 정비 로드맵 마련.”
- xxxi PIPC (2023). “전면 개정 개인정보 보호법, 9월 15일 시행.” Available at: <https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=9145>
- xxxi ZDNet Korea (2023). “4기 AI 법제정비단 발족...연내 로드맵2.0 발표.” Available at: <https://zdnet.co.kr/view/?no=20230816143413>
- xxxi Korea National Assembly (2023). Legislative bill on fostering artificial intelligence and creating a foundation of trust.
- xxxi Korea National Assembly (2023). [2120898] 채용절차의 공정화에 관한 법률 일부개정법률안(박대수의원 등 15인). Available at: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_B2A3H0G3F2B3A0X9F1E2O0N4M1I2H2; Korea National Assembly (2023). [2121075] 채용절차의 공정화에 관한 법률 일부개정법률안(신영대의원 등 10인). Available at: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_Y2X3Y0G3F0D6C1Y7X3F7N5L5K6I6F9; Korea National Assembly (2023). [2120799] 채용절차의 공정화에 관한 법률 일부개정법률안(한준호의원 등 17인). Available at: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_M2L3K0R3Q0R9Q1Q4Z2X8W0X8V5K7J4; Korea National Assembly (2023). [2122180] 콘텐츠산업 진흥법 일부개정법률안(이상현의원 등 10인). Available at: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_R2P3Q0X5Y1X7X1V3W4V4C5D6B5C8A2; Korea National Assembly (2023). [2122569] 공직선거법 일부개정법률안(송석준의원 등 12인). Available at: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_D2B3C0K5K2J6H1I2G0H4O1P6N2O0M4; Korea National Assembly (2023). [2120001] 정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안(김홍걸의원등 13인). Available at: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_N2S3F0B1O1Q2K1V1E0B9U4B4V0S3U0
- xxxi ICT Regulatory Sandbox (2023). “Demonstration Tests.” Available at: <https://www.sandbox.or.kr/apply/regulation.do>
- xxxi Ibid.
- xxxi European Parliament (2023). “Texts adopted: Artificial Intelligence Act.” Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html
- xi Ministry of Enterprises and Made in Italy (2023). “Italy, Germany and France agree on strengthening their cooperation on Artificial Intelligence.” Available at: <https://www.mimit.gov.it/en/media-tools/news/italy-germany-and-france-agree-on-strengthening-their-cooperation-on-artificial-intelligence>
- xli The White House (2023). “Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.”
- xlii Skadden, Arps, Slate, Meagher & Flom LLP and Affiliates (2023). “AI Insights: Biden Administration Passes Sweeping Executive Order on Artificial Intelligence.” Available at: <https://www.skadden.com/insights/publications/2023/11/biden-administration-passes-sweeping-executive-order-on-artificial-intelligence>
- xliii The White House (2023). “Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.”
- xliv OECD.AI (2023). “Shaping an enabling policy environment for AI (Principle 2.3).” Available at: <https://oecd.ai/en/dashboards/ai-principles/P12>
- xlv European Parliament (2023). “Texts adopted: Artificial Intelligence Act.” Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html
- xlvi Korea National Assembly (2023). AI Responsibility and Regulation Bill. Available at: http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_W2E3F0D8D0C1A0B9I3J1I2I1G4F0G1
- xlvii Davis Wright Tremaine LLP (2023). “NIST Releases Final Risk Management Framework for Developing Trustworthy AI.” Available at: <https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2023/01/ai-risk-management-framework-nist>
- xlviii Microsoft (2023). Governing AI: A Blueprint for the Future. Available at: <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw>
- xlix Microsoft (2017). “The need for a Digital Geneva Convention.” Available at: <https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/>
- i Microsoft (2018). “Facial recognition technology: The need for public regulation and corporate responsibility.” Available at: <https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/>
- ii Microsoft (2020). “Finally, progress on regulating facial recognition.” Available at: <https://blogs.microsoft.com/on-the-issues/2020/03/31/washington-facial-recognition-legislation/>
- iii Microsoft (2023). “Microsoft announces new Copilot Copyright Commitment for customers.” Available at: <https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/>
- iiii Digital Public Sphere (2023). “Digital Bill of Rights.” Available at: http://beingdigital.kr/front/digital_rights.do
- liv MSIT (2020). “과기정통부, 인공지능 윤리기준 마련.” Available at: <https://www.msit.go.kr/bbs/view.do?sCode=user&mPid=112&mlId=113&bbsSeqNo=94&nttSeqNo=3179742>
- lv MSIT (2020). “인공지능 시대를 준비하는 법·제도·규제 정비 로드맵 마련.” Available at: <https://www.msit.go.kr/bbs/view.do?sCode=user&mPid=112&mlId=113&bbsSeqNo=94&nttSeqNo=3179749>
- lvi MSIT (2023). “South Korea presents a new digital order to the world!” Available at: <https://www.korea.net/Government/Briefing-Room/Press-Releases/view?articleId=7042&insttCode=A110439&type=O>
- lvii MSIT (2023). “디지털 접근성 컨퍼런스 개최.” Available at: <https://www.msit.go.kr/bbs/view.do?sCode=user&nttSeqNo=3183225&bbsSeqNo=94>
- lviii Digital Government (n.d.). Digital Government Masterplan 2021-2025. Available at: <https://www.dgovkorea.go.kr/>
- lix AI Times (2023). “NIA, ‘디지털포용 포럼’ 운영위원회 개최...”소외없는 디지털 사회를 위해.” Available at: <https://www.aitimes.com/news/articleView.html?idxno=155154>
- lx OECD AI Policy Observatory (2023). OECD AI Principles overview. Available at: <https://oecd.ai/en/ai-principles>
- lxi NIST (2023). AI Risk Management Framework. Available at: <https://www.nist.gov/itl/ai-risk-management-framework>
- lxii The White House (2023). “Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.” Available at: <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>

References

- lxiii** GOV.UK (2023). Chair’s Summary of the AI Safety Summit 2023, Bletchley Park. Available at: <https://www.gov.uk/government/publications/ai-safety-summit-2023-chairs-statement-2-november/chairs-summary-of-the-ai-safety-summit-2023-bletchley-park>
- lxiv** Microsoft (2018). “AI for Accessibility, a \$25 million program to empower people with disabilities.” Available at: <https://blogs.microsoft.com/accessibility/ai-for-accessibility/>
- lxv** Microsoft (2021). “Our new accessibility commitment.” Available at: <https://news.microsoft.com/accessibility-commitment/>
- lxvi** Microsoft (2022). “Because every vote counts: making elections more accessible.” Available at: <https://blogs.microsoft.com/on-the-issues/2022/09/12/disability-voting-rights-week-index/>
- lxvii** DE News (2020). “MS-SK Telecom-Korea Employment Agency for Persons with Disabilities, ‘AI for Accessibility’ Seminar.” Available at: <https://www.denews.co.kr/news/articleView.html?idxno=16576>
- lxviii** NEWSBYTES.PH (2023). “Microsoft enabler program helped 350 Pinoy PWDs get employed.” Available at: <https://newsbytes.ph/2023/05/31/microsoft-enabler-program-helped-350-pinoy-pwds-get-employed/>
- lxix** Microsoft (n.d.). Our Accessibility Evolution Model. Available at: <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4wNu4>
- lxx** Digital Public Sphere (2023). “Digital Bill of Rights.”
- lxxi** KISDI (2022). 인공지능 윤리기준 실천을 위한 자율점검표. Available at: <https://www.kisdi.re.kr/cmm/fileDown.do?key=170565&type=bbs>; TTA (2022). 신뢰할 수 있는 인공지능 개발안내서. Available at: <https://tta.or.kr/tta/downloadBbsFile.do?atchmfnNo=9722>; KISDI (2022). 인공지능 윤리교육 교재. Available at: <https://www.kisdi.re.kr/bbs/view.do?bbsSn=113985&key=m2101113056011&pageIndex=1&sc=&sw=>
- lxxii** KISDI (2023). “2023 인공지능 윤리기준 실천을 위한 자율점검표(안) 공개.” Available at: <https://www.kisdi.re.kr/bbs/view.do?bbsSn=114068&key=m2101113055944>
- lxxiii** Korea Policy Briefing (2022). “인공지능, 교육현장에서 안전하게 활용해요!” Available at: <https://www.korea.kr/briefing/pressReleaseView.do?newsId=156520749>
- lxxiv** KDI (2023). “AI 디지털교과서로 1:1 맞춤 교육시대 연다.” Available at: <https://eiec.kdi.re.kr/policy/materialView.do?num=239513&pg=&pp=20&topic=O>
- lxxv** NIA (2023). “About NIA.” Available at: https://eng.nia.or.kr/site/nia_eng/04/1040100000002016093002.jsp
- lxxvi** OECD AI Policy Observatory (2023). OECD AI Principles overview.
- lxxvii** OECD AI Policy Observatory (2023). “How countries are implementing the OECD Principles for Trustworthy AI.” Available at: <https://oecd.ai/en/wonk/national-policies-2>
- lxxviii** European Commission (2023). Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems. Available at: <https://ec.europa.eu/newsroom/dae/redirection/document/99641>
- lxxix** The White House (2023). “Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.”
- lxxx** EY (2023). Key takeaways from the Biden administration executive order on AI. Available at: https://www.ey.com/en_us/public-policy/key-takeaways-from-the-biden-administration-executive-order-on-ai
- lxxx i** GOV.UK (2023). Chair’s Summary of the AI Safety Summit 2023, Bletchley Park.
- lxxxii** Hankookilbo (2023). “MS AI School changed my values... Realize the ‘joy of learning.’” Available at: <https://www.hankookilbo.com/News/Read/A2023062010290004218>
- lxxxiii** Venturesquare (2023). “Jump up together with digital education without discrimination!” Available at: <https://www.venturesquare.net/890280>
- lxxxiv** Microsoft and KCCI (2023). The Economic Impact of Generative AI: The Future of Work in South Korea.
- lxxxv** Digital Public Sphere (2023). “Digital Bill of Rights.”
- lxxxvi** AI for Good (2023). “The Ministry of Science and ICT (MSIT) of the Republic of Korea partners with the ITU to foster a hospitable environment for artificial intelligence (AI).” Available at: <https://aiforgood.itu.int/the-ministry-of-science-and-ict-partners-with-the-itu-to-foster-a-hospitable-environment-for-artificial-intelligence-ai/>
- lxxxvii** Korea IT News (2023). “디지털 ESG 협의체’ 출범.” Available at: <http://www.koit.co.kr/news/articleView.html?idxno=116451>
- lxxxviii** MSIT (2023). “South Korea presents a new digital order to the world!”
- lxxxix** Korea Policy Briefing (2023). “대한민국 인공지능 도약방안 발표.” Available at: <https://www.korea.kr/briefing/policyBriefingView.do?newsId=156590066&pWise=sub&pWiseSub=C3>
- xc** OECD.AI (2023). “Inclusive growth, sustainable development and well-being (Principle 1.1).” Available at: <https://oecd.ai/en/dashboards/ai-principles/P5>
- xc i** European Commission (2023). “Hiroshima Process International Guiding Principles for Advanced AI system.” Available at: <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system>
- xcii** The White House (2023). “Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.”
- xciii** GOV.UK (2023). Chair’s Summary of the AI Safety Summit 2023, Bletchley Park.
- xciv** Microsoft (2020). Microsoft will be carbon negative by 2030. Available at: <https://blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/>
- xcv** Microsoft (n.d.). A planetary computer for a sustainable future. Available at: <https://planetarycomputer.microsoft.com/>
- xcvi** Microsoft (2022). Environmental Sustainability Report. Available at: 2022 Environmental Sustainability Report (microsoft.com)
- xcvii** Microsoft (2021). This non-profit is protecting vulnerable communities from the effects of climate change with AI. Available at: <https://news.microsoft.com/source/asia/features/seeds-india-climate-change-ai/>
- xcviii** Microsoft (2023). Accelerating Sustainability with AI: A Playbook. Available at: <https://blogs.microsoft.com/on-the-issues/2023/11/16/accelerating-sustainability-ai-playbook/>
- xcix** Microsoft (2023). The 2023 Impact Summary. Available at: The 2023 Impact Summary-Building a responsible future (microsoft.com)
- c** Law Times (2020). “과학기술정보통신부 국가 인공지능 윤리기준 발표.”
- ci** KISDI (2022). 인공지능 윤리기준 실천을 위한 자율점검표.; TTA (2022). 신뢰할 수 있는 인공지능 개발안내서.; KISDI (2022). 인공지능 윤리교육 교재.
- cii** KISDI (2023). “「2023 인공지능 윤리기준 실천을 위한 자율점검표(안)」 공개.”
- ciii** KDI (2023). Digital Bill of Rights.
- civ** ZDNet Korea (2023). “4기 AI 법제정비단 발족…연내 로드맵2.0 발표.”
- cv** KDI (2023). “과기정통부, ‘인공지능 윤리·신뢰성 확보 추진계획’ 발표.” Available at: <https://eiec.kdi.re.kr/policy/callDownload.do?num=243950&filenum=1&dtim=20231025165822>
- cv i** NIA (2023). “About NIA.”
- cvii** Korea National Assembly (2023). Legislative bill on fostering artificial intelligence and creating a foundation of trust.
- cviii** Fortune Korea (2023). “약일까 독일까…인공지능(AI)법의 운명은.” Available at: <https://www.fortunekorea.co.kr/news/articleView.html?idxno=29373>
- cix** Ibid.
- cx** Korea National Assembly (2023). AI Responsibility and Regulation Bill.
- cx i** PIPC (2023). “전면 개정 개인정보 보호법, 9월 15일 시행.”
- cxii** Korea National Assembly (2023). [2120130] 개인정보 보호법 일부개정법률안(김영배의원 등 12인).
- cxiii** Korea National Assembly (2023). [2120898] 채용절차의 공정화에 관한 법률 일부개정법률안(박대수의원 등 15인).; Korea National Assembly (2023). [2121075] 채용절차의 공정화에 관한 법률 일부개정법률안(신영대의원 등 10인).; Korea National Assembly (2023). [2120799] 채용절차의 공정화에 관한 법률 일부개정법률안(한준호의원 등 17인).
- cxiv** Korea National Assembly (2023). [2122180] 콘텐츠산업 진흥법 일부개정법률안(이상현의원 등 10인).
- cxv** Korea National Assembly (2023). [2122537] 저작권법 일부개정법률안(황보승희의원 등 10인).

References

- dxix Ibid
- dxxi EY (2023). The EU AI Act: What it means for your business. Available at: https://www.ey.com/en_ch/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business
- dxxii European Parliament (2023). Briefing on EU Legislation in Progress, Artificial Intelligence Act.
- dxxiii Centre for Security and Emerging Technology (2023). AI Act: A Primer. Available at <https://cset.georgetown.edu/article/the-eu-ai-act-a-prime>
- dxxiv European Parliament (2023). Briefing on EU Legislation in Progress, Artificial Intelligence Act.
- dx xv Ibid.
- dx xvi EY (2023). The EU AI Act: What it means for your business.
- dx xvii Centre for Security and Emerging Technology (2023). AI Act: A Primer.
- dx xviii European Parliament (2023). Briefing on EU Legislation in Progress, Artificial Intelligence Act.
- dx xix Ibid.
- dx xx Centre for Security and Emerging Technology (2023). AI Act: A Primer.
- dx xxi European Parliament (2023) Briefing on EU Legislation in Progress, Artificial Intelligence Act.
- dx xxi Ibid.
- dx xxiii European Parliament (2023). Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI. Available at: <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>
- dx xxiv European Parliament (2023). EU AI Act: first regulation on artificial intelligence. Available at: <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- dx xxv The White House (2023). “Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI.”
- dx xxvi Reuters (2023). “Adobe, others join voluntary US scheme to manage AI risks.”
- dx xxvii GOV.UK (2023). “Safety Testing: Chair’s Statement of Session Outcomes, 2 November 2023.” Available at: <https://www.gov.uk/government/publications/ai-safety-summit-2023-chairs-statement-safety-testing-2-november/safety-testing-chairs-statement-of-session-outcomes-2-november-2023>
- dx xxviii Ibid.
- dx xxix Microsoft (2023). Transparency Note for Azure OpenAI Service. Available at: <https://learn.microsoft.com/en-us/legal/cognitive-services/openai/transparency-note?tabs=text>
- cx c Microsoft (2023). Empowering responsible AI practices. Available at: <https://www.microsoft.com/en-us/ai/responsible-ai>
- cx ci Microsoft (2023). Copilot Copyright Commitments. Available at: <https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/>
- cx cii Microsoft (2023). Product Terms. Available at: <https://www.microsoft.com/licensing/terms/product/ForOnlineServices>
- cx ciii Hankookilbo (2023). “MS AI School changed my values... Realize the ‘joy of learning.’”
- cx civ Microsoft (2019). “Protecting the world’s languages for an inclusive future.” Available at: <https://news.microsoft.com/apac/2019/01/03/protecting-the-worlds-languages-for-an-inclusive-future/>
- cx cv Microsoft (2021). “Bentley Systems case study.” Available at: <https://partner.microsoft.com/en-rs/case-studies/bentley-systems>
- cx cvi Microsoft (2023). “Microsoft Korea helps Lotte Hotel build a hyper-automated work environment with Power Platform.” Available at: https://news.microsoft.com/ko-kr/2023/07/10/lottehotel_powerplatform/
- cx cvii Microsoft (2019). “Fighting fire with AI.” Available at: <https://news.microsoft.com/apac/features/fighting-fire-with-ai/>