RESEARCH ON AI POLICY AND ISSUES IN KEY AREAS IN SOUTH KOREA

PUBLIC SECTOR
LAW ENFORCEMENT
EDUCATION and
SOCIAL WELFARE

APC
ASSOCIATION FOR PROGRESSIVE COMMUNICATIONS

진보네트워크센터
Korean Progressive Network 'Jinbonet'

정보인권연구소
Institute for Digital Rights

Research on AI Policy and Issues in key Areas in South Korea
: Public Sector, Law Enforcement, Education and Social Welfare

Published on 15 February 2025
This report was produced by the Korean Progressive Network Jinbonet and Institute for Digital Rights
with support from the Association for Progressive Communications (APC).

The following authors participated in the research and writing of this report:
- Yeo-kyung Chang (Executive Director, Institute for Digital Rights)
- Byeong-rok Lee (Director of Information and Communications, Korean Teachers and Education Workers Union)
- Ji-hee Kim (Chairman of the Board, Nurimashil Friends Cooperative)
- Byoung-il Oh (Director, Korean Progressive Network Jinbonet)
- Heewoo (Activist, Korean Progressive Network Jinbonet)

Any enquiries regarding this publication should be sent to:
Byoung-il Oh, antiropy@gmail.com
The link to the Korean version of the report.:
https://act.jinbo.net/wp/50224/
Address:
23, Dongnimmun-ro 8-gil, Seodaemun-gu, Seoul, Republic of Korea 03745

Table of Contents

Executive Summary

Increasingly diverse AI systems are being rapidly implemented across various sectors of society. As a result, several issues are emerging, including AI systems' instability, opacity, bias, and misuse for surveillance purposes.

In South Korea, civil society has raised various concerns surrounding AI systems. These include hate speech and privacy violations by the AI chatbot "Lee Ruda," the Incheon Airport Immigration Control System Upgrade Project that provided facial recognition data to AI developers without data subjects' consent, AI recruitment systems being implemented in public institutions without proper risk and performance assessments, and the Ministry of Education's forceful pursuit of AI digital textbooks without adequate preparation. However, additional research is needed on what types of AI systems are being introduced across Korean society and under what procedures and policies. This report serves this purpose. It analyzes the standards for AI regulation, implementation status, and actual or potential problems in key areas including public administration, law enforcement, education, and social welfare.

Chapter 1 analyzes the current state of AI implementation and regulatory framework in South Korea's public sector. While various AI systems such as chatbots, security monitoring, and administrative automation are being deployed across public institutions, there is a lack of integrated management systems and clear guidelines. The article identifies key areas for improvement including the introduction of an AI registration system, conducting human rights impact assessments, and securing AI expertise. It particularly emphasizes the need for stricter regulation and management of public sector AI, given its direct impact on citizens' rights and obligations.

Chapter 2 examines the current state of AI implementation in South Korean law enforcement, particularly focusing on the police and immigration authorities. The police are actively developing and deploying various AI systems, including intelligent CCTV, crime prediction, real-time behavioral analysis, and automated tracking systems, as part of their comprehensive Public Security Technology Plan. However, the report raises significant concerns about human rights implications, including lack of transparency, excessive personal data collection, real-time surveillance capabilities, and insufficient legal frameworks and oversight mechanisms. A key example is the controversial Immigration AI project, which used massive amounts of biometric data without proper consent or legal basis, highlighting the broader challenges of balancing law enforcement capabilities with privacy rights and civil liberties.

Chapter 3 examines the current state of AI implementation in South Korea's education sector, with a particular focus on the controversial AI Digital Textbook (AIDT) initiative planned for 2025. While the Ministry of Education is pushing forward with

implementing AIDT as a key tool for personalized learning, the initiative faces significant opposition from teachers, parents, and civil society organizations. Major concerns include insufficient stakeholder consultation, questionable educational effectiveness, potential privacy issues related to extensive student data collection, and the substantial financial burden on local education offices.

Chapter 4 examines the implementation of AI in South Korea's social welfare sector, focusing on how artificial intelligence is being used to provide services to vulnerable populations. While the government is promoting data-driven welfare through AI systems for health monitoring, fraud detection, and welfare recipient identification, there are significant concerns about privacy protection, data consent practices, and potential negative social impacts. The article particularly highlights issues surrounding the AI and IoT-based Senior Health Management Project and welfare fraud detection systems, emphasizing the need for proper regulations and safeguards.

Along with analyzing the status of AI implementation in key sectors, Chapter 5 summarizes the controversy surrounding the establishment of Korea's AI Framework Act. As international awareness of AI risks grows, various countries are discussing the introduction of AI regulations, such as the EU AI Act. In South Korea, the AI Framework Act has been a subject of social controversy over the past few years, and finally, on December 26, 2024, an AI Framework Act focusing on industrial development passed the National Assembly. Key issues include the lack of provisions for prohibited AI systems, narrow scope of high-impact AI regulation, insufficient penalties for violations, inadequate rights and remedies for affected persons, and the controversial exemption of defense and national security AI systems from regulation.

We hope this report will help develop international norms to protect citizens' safety and human rights from the risks of AI. The Korean Progressive Network Jinbonet plans to continue publishing reports on key issues and developments surrounding AI in South Korea.

Chapter 1

Public Sector AI: Regulatory Framework and Current State

1.1. Introduction

Various AI systems are already being implemented and utilized across the public sector, including central administrative agencies, local governments, and public institutions. These include chatbots for public services, digital evidence analysis systems for investigative agencies, environmental monitoring AI, and other diverse AI systems. Furthermore, more advanced AI systems are being rapidly introduced. However, because there is no integrated management of implemented AI systems, it is not easy to accurately identify what types of AI systems with what functions have been introduced in the public sector for what purposes. Additionally, there is insufficient guidance on what principles and procedures should be followed when developing or procuring AI systems. AI systems introduced in the public sector often affect citizens' rights and obligations, and unlike private companies' AI systems, citizens have no alternative choices. Therefore, policies and supervision systems that can strictly regulate public sector AI systems need to be established.

1.2. AI-Related Regulations in the Public Sector

1.2.1. Electronic Government Act and its Enforcement Decree, and Guidelines for Implementation and Management of Intelligent E-Government Services

Article 18-2, newly established in the E-Government Act amended on June 8, 2021, enables the use of AI and other technologies in providing e-government services and allows the Minister of the Interior and Safety to provide necessary administrative, financial, and technical support. However, while it provides a basis for utilizing AI technology in e-government services, it does not specifically regulate principles or procedures for introduction of AI systems considering AI's characteristics and risks. However, while the Minister of the Interior and Safety is required to announce detailed matters necessary for the selection and management of supported projects, this is outlined in the 'Guidelines for the Implementation and Management of Intelligent E-Government Services.' Article 7 of these guidelines stipulates that when selecting support projects, factors such as 'the concrete details of project objectives and implementation plans for AI technology adoption' and 'the sustainability of securing AI training data and the appropriateness of

its quality' should be considered. Nevertheless, these are merely considerations for selecting support projects and not evaluation criteria for the AI systems that are actually planned for adoption. Article 8 of these guidelines only stipulates that administrative agencies should "strive" to "comply with prior notification obligations in decision-making processes and guarantee rights to refuse service, raise objections, and demand explanations" when providing AI-enabled services. Furthermore, the E-Government Act and guidelines apply only to e-government services, not to all AI systems used by public institutions for business purposes.

Electronic Government Act

Article 18-2 (Provision of Intelligent Electronic Government Services) (1) The head of an administrative agency, etc. may provide electronic government services by utilizing technologies, such as artificial intelligence.
(2) The Minister of the Interior and Safety may provide administrative, financial, technical, and other necessary support to help the head of an administrative agency, etc. efficiently utilize technologies, such as artificial intelligence.
(3) Types of technologies, such as artificial intelligence under paragraphs (1) and (2) and matters necessary for the utilization and support thereof shall be prescribed by the National Assembly Regulations, the Supreme Court Regulations, the Constitutional Court Regulations, the National Election Commission Regulations, or by Presidential Decree.
[This Article Newly Inserted on Jun. 8, 2021]

Enforcement Decree of the Electronic Government Act

Article 15-2 (Introduction and Utilization of Intelligent Electronic Government Services) (1) Technologies such as artificial intelligence, etc. that can be utilized for the provision of electronic government services pursuant to Article 18-2 (1) of the Act shall be as follows:
1. Natural language processing (referring to a technology that analyzes and processes human language using computers);
2. Voice recognition;
3. Video recognition;
4. Other technologies necessary for providing intelligent electronic government services, which realize learning, reasoning, judgment, etc. by electronic means.
(2) The Minister of the Interior and Safety may support the following projects pursuant to Article 18-2 (2) of the Act:
1. Projects for applying and demonstrating artificial intelligence, etc. technologies to electronic government services;
2. Projects of establishing common foundation for utilizing artificial intelligence, etc. technologies for various electronic government services;
3. Projects of converging artificial intelligence, etc. technologies with other technologies or services, such as big data analysis techniques;
4. Other projects necessary for introducing or utilizing intelligent electronic government services.
(3) Detailed matters necessary for the selection, management, etc. of projects under paragraph (2) shall be determined and publicly notified by the Minister of the Interior and Safety.
[This Article Newly Inserted on Dec. 9, 2021]

Guidelines for Implementation and Management of Intelligent E-Government Services

Article 7 (Selection and Management of Intelligent E-Government Service Support Projects)

② When selecting intelligent e-government service support projects, the Minister of the Interior and Safety shall consider the following matters:

1. Feasibility and efficiency of the project
2. Concrete details of project objectives and implementation plans for AI technology adoption
3. Sustainability of securing AI training data and appropriateness of its quality
4. Possibility of public administration automation and provision of new high-quality intelligent public services
5. Other matters determined by the Minister of the Interior and Safety

Article 8 (Provision and Management of Intelligent E-Government Services)

① When providing intelligent e-government services using AI technology, the heads of administrative agencies shall comply with prior notification obligations in the decision-making process and strive to guarantee the rights to refuse service use, raise objections, and demand explanations.

② The heads of administrative agencies shall strive to prevent the negative functions of AI utilization and implement human-centered AI when introducing and operating intelligent e-government services.

③ The heads of administrative agencies shall establish security measures necessary for providing secure intelligent e-government services.

1.2.2 Framework Act on Intelligence Informatization

The Framework Act on Intelligence Informatization lists various intelligent information technologies, and while it does not use the term "artificial intelligence," the technology described in Article 2, Paragraph 4, Item (a) can be considered AI-related. Regarding public institutions' use of AI, Article 14 only stipulates that public intelligent informatization should be promoted and necessary measures should be prepared for its efficient implementation. Article 46, though not limited to AI systems or public institutions, requires intelligent information service providers to “endeavor to improve access and user convenience for the disabled and elderly” when providing services. It also requires "national agencies, etc. to establish policy measures necessary to promote the preferential purchase of intelligent information products which guarantee access to information for the disabled and elderly, and user convenience." Article 57 requires national agencies and local governments to establish information protection measures when providing and using intelligent information services. Article 60 stipulates that the safety protection measures' content and methods for intelligent information technology and services should be determined and announced, but it appears that these have not yet been announced.

Framework Act on Intelligence Informatization

Article 2 (Definitions)

4. The term “intelligent information technology” means any of the following technologies or technology that converges and applies such technologies:

(a) Technology that electronically realizes learning, reasoning and decision-makings;

(b) Technology that electronically collects, analyzes and processes data (referring to all kinds of data or knowledge expressed in codes, letters, voice, sound, image, etc.);

Article 14 (Promotion of Intelligent Informatization of Public Sector) (1) National agencies and similar entities shall promote intelligent informatization of the functions under their jurisdiction, including administration, public health, social welfare, education, culture, the environment, transportation, logistics, science and technology, disasters and safety, security, national defense and energy (hereinafter referred to as “intelligent informatization of the public sector”) for such purposes as furthering intelligent informatization of public services and increasing the convenience of citizens.

Article 46 (Guaranteeing Access to and Use of Information by Persons with Disabilities and Senior Citizens)

(2) Intelligent information services providers shall endeavor to improve access by persons with disabilities, senior citizens, etc. and user convenience when providing such services.

(4) In purchasing intelligent information products, national agencies and similar entities shall establish policy measures necessary to promote the preferential purchase of intelligent information products which guarantee access to information by persons with disabilities, senior citizens, etc. and user convenience.

Article 57 (Establishment of Policy Measures on Protection of Information) (1) National agencies and local governments shall establish policy measures to protect information through the entire course in which information is processed or intelligent information services are provided or used.

Article 60 (Safety Protection Measures) (1) The Minister of Science and ICT may determine and publicly notify the details and methods of minimum necessary protection measures as specified in the following in consultation with the heads of relevant agencies, including the Minister of the Interior and Safety, in order to ensure the safety of intelligent information technology and intelligent information services:

1. Matters concerning prevention of malfunctions of intelligent information technology and intelligent information services;
2. Matters concerning prevention of electronic intrusions, such as unauthorized access to and manipulation of intelligent information technology and intelligent information services;
3. Matters concerning storage, management, provision, etc. of the access logs, operation and use logs of intelligent information technology and intelligent information services;
4. Matters concerning shutting down (hereinafter referred to as “emergency shutdown”) the operation of intelligent information technology and provision of intelligent information services externally in an emergency situation and provision of algorithm necessary for emergency shutdown;
5. Other matters necessary to ensure the safety of intelligent information technology and intelligent information services.

1.2.3. General Act on Public Administration, Article 20

Article 20 of the General Act on Public Administration allows for fully automated systems, including AI systems, to make administrative dispositions when there is no discretion involved. When AI-based administrative dispositions are made automatically without human intervention at any stage, issues of democratic legitimacy may arise. Article 20 was legislated to promote digitalization of administration and establish legislative standards for automatic dispositions to enhance administrative efficiency and public convenience. To prevent indiscriminate automatic dispositions, legal grounds are required for their introduction, and automatic dispositions are not allowed in cases where discretion exists (according to Commentary on the General Act on Public Administration).

General Act on Public Administration
Article 20 (Automatic Dispositions) An administrative authority may impose a disposition using a fully-automated system (including systems in which artificial intelligence technologies are employed): Provided, That the same shall not apply to dispositions imposed at its discretion.

However, the procedural rights of those subject to administrative dispositions should not be weakened or excluded in laws stipulating automatic dispositions. With only this provision, it remains unclear what safety measures are necessary for automated dispositions and how they will affect the rights of the parties subject to these dispositions. Although this provision was implemented in March 2021, as of late 2024, only Article 20-2 (Automation of Import Declaration Acceptance) of the 'Special Act on Imported Food Safety Management' contains provisions for automatic disposition. Even in this law, detailed matters such as scope and procedures are to be determined by the enforcement decree. Meanwhile, in response to inquiries by the author, the Ministry of Government Legislation has stated that they are establishing legislative standards that administrative agencies must follow when introducing automatic dispositions, and plan to distribute 'Legislative Guidelines for Automatic Dispositions' to central administrative agencies, etc. in 2024.

Furthermore, Article 37-2 of the Personal Information Protection Act grants data subjects the right to refuse decisions and demand explanations when decisions made through fully automated systems (including AI systems) processing personal information significantly affect their rights or obligations. However, this excludes automatic dispositions under Article 29 of the General Act on Public Administration. This exclusion is reportedly based on the consideration that the rights of data subjects are already guaranteed under general provisions for administrative dispositions, such as the Administrative Procedures Act and Administrative Litigation Act. Additionally, for automated decisions in the public sector that are not automatic dispositions under Article 20 of the General Act on Public Administration, the provisions on automated decisions in the Personal Information Protection Act apply, thus requiring the protection of data subjects' rights regarding automated decisions.

1.2.4. Draft Practitioner's Guide for AI Implementation in the Public Sector

In April 2021, the Public Intelligence Policy Division of the Digital Information Bureau, Ministry of the Interior and Safety, released a <Draft Practitioner's Guide for AI Implementation in the Public Sector>. While this draft guide was explicitly stated as 'a preliminary version for gathering opinions' and subject to content changes, it had not been adopted as official guidelines by the end of 2024. When questioned about the reason, the Ministry responded that they are "currently developing a 'Public AI Implementation and Utilization Strategy' to systematically expand AI adoption in the public sector, and plan to establish necessary guidelines accordingly."

This draft guide presents procedures differentiated from general informatization projects, considering AI's unique characteristics. For example, in the planning stage, it requires establishing data collection and procurement plans and conducting AI ethics reviews. In the development stage, it calls for selecting AI models and algorithms, evaluating learning performance, and establishing data management plans. For the operational stage, it recommends monitoring AI dysfunction and establishing retraining plans.

The draft guide recognizes the challenges faced by AI project managers in the public sector , including ▲difficulties in vendor selection and project inspection due to the lack of common standards for AI performance evaluation, and ▲limitations in accessing internal structures and handling unexpected situations due to undisclosed algorithms.

It is questionable why official and systematic guidelines have not been published, despite various AI system implementations being pursued in the public sector since this draft guide was created.

1.2.5. ChatGPT Usage Guide

The first ChatGPT usage guidelines were published amid a lack of formal guidance on AI implementation and use in the public sector. This initiative appears to have been triggered by President Yoon Suk-yeol's directive in January 2023, instructing presidential staff and government departments to actively utilize ChatGPT in their work.

In May 2023, the Ministry of the Interior and Safety released a <Guide to ChatGPT Usage and Precautions> for government officials. This concise 8-page manual outlines three key areas for public sector application: ▲information research capabilities, ▲language skills, and ▲computer-related tasks, complete with seven detailed use cases and examples. The guide also addresses ChatGPT's limitations, warning against inputting incomplete decision-making matters, confidential information, or personal data. It emphasizes the critical importance of fact-checking ChatGPT's responses.

The National Intelligence Service (NIS) followed suit in June 2023 by publishing <Security Guidelines for Generative AI Including ChatGPT>, as the agency responsible for public sector cybersecurity. The NIS outlined seven major security risks associated with generative AI technology and provided safety protocols for its use in the public sector. The guidelines also include security measures and key considerations for public institutions implementing AI

systems, including generative AI. However, as public institutions continue to adopt various AI systems with different functions and features, there's a growing need for comprehensive guidelines establishing clear principles and procedures for AI implementation. The Seoul Digital Foundation also contributed to this knowledge base by publishing <ChatGPT Use Cases and Tips - Workplace Edition> in March 2023.

Questions have been raised about the wisdom of this approach - specifically, whether it's appropriate for the president to mandate the use of a specific AI application without first establishing fundamental principles and procedures for AI adoption in the public sector, and for government departments to hastily issue guidelines based solely on presidential directives.

1.2.6. Guidelines for Implementing and Using Advanced AI in the Public Sector

In April 2024, the Presidential Committee on Digital Platform Government and the National Information Society Agency (NIA) published <Guidelines for Implementing and Using Advanced AI in the Public Sector>. These guidelines aim to provide public institutions with standards, procedures, and key considerations for adopting and utilizing advanced AI systems.

The guidelines outline five core principles for implementing advanced AI:
• Timely adoption and utilization of cutting-edge private sector technology
• Simultaneous innovation of administrative processes and organizational culture
• Breaking down inter-departmental barriers to create a unified government
• Ensuring national security and protecting citizens' rights
• Compliance with AI ethical standards announced by the Ministry of Science and ICT in December 2020

The implementation process consists of six key stages:
1. Data security level assessment
2. Cloud infrastructure planning: private vs. public cloud options
3. Data training methodology: general-purpose LLMs, fine-tuning, post-training, etc.
4. Service implementation approach: service purchase or procurement through bidding
5. Service level objectives: accuracy, response time, availability, etc.
6. Maintenance and operations

While this represents the first official guidelines for AI implementation in the public sector, it specifically focuses on advanced AI systems. Although the creation of general guidelines is a positive development, they remain at a somewhat introductory level. Furthermore, while the principles emphasize protecting citizens' rights and adhering to AI ethical standards, and preliminary considerations include understanding and preventing risks associated with generative AI, these aspects are not explicitly incorporated into the implementation procedures. The guidelines also lack detailed explanations of how to identify and mitigate specific risks.

1.2.7. AI-Related Ordinances in Local Governments

Several local governments have been establishing their own AI ordinances or regulations for fostering and supporting the AI industry. For instance, Gyeonggi Province and Bucheon City have enacted basic AI ordinances, while Gyeonggi Province, South Gyeongsang Province, North Gyeongsang Province, Gwangju Metropolitan City, Daegu Metropolitan City, and Sejong Special Self-Governing City have implemented ordinances for fostering and supporting the AI industry. Gyeonggi Province, in addition to the two ordinances mentioned above, has also established ordinances for fostering AI startups and creating an AI ethical foundation.

However, these local government ordinances have limitations. Their jurisdiction is restricted to their respective regions, and the regulatory content tends to lack specificity. For example, while the Gyeonggi Province Basic AI Ordinance includes definitions for prohibited AI and high-risk AI systems, it only broadly states that prohibited AI development is forbidden in principle and that high-risk AI systems are permitted within the strict application of relevant laws and regulations. The ordinance fails to specify which AI systems fall under the prohibited or high-risk categories or detail the specific regulatory requirements. Moreover, most other ordinances primarily focus on fostering and supporting the AI industry rather than addressing comprehensive regulatory frameworks.

Gyeonggi Province Basic AI Ordinance

Article 2 (Definitions)

The terms used in this ordinance are defined as follows:

1. "Artificial Intelligence" refers to software, computer systems, or other devices designed to operate with varying levels of autonomy, performing functions characteristic of human intelligence such as learning, reasoning, perception, and natural language understanding.
2. "Prohibited AI" refers to AI systems that violate relevant laws and social norms, particularly those clearly considered to threaten human dignity, life, liberty, and equality.
3. "High-Risk AI" refers to AI systems used in areas where there is significant potential impact on human life, physical safety, and fundamental rights.
4. "Low-Risk AI" refers to all AI systems that do not fall under the categories of prohibited AI or high-risk AI.

Article 3 (Basic Principles)

The development and use of AI must adhere to the following basic principles:

1. AI must be developed and used for the advancement and convenience of humanity.
2. The development and use of AI must not discriminate against any individual or group based on gender, age, race, ethnicity, region, physical condition, religion, economic circumstances, or political views. It must ensure accessibility for socially disadvantaged and vulnerable groups.
3. The development and use of AI must guarantee individuals' right to personal information self-determination and ensure reliability and transparency.
4. The development of prohibited AI shall be forbidden in principle; high-risk AI shall be permitted only within the strict application of relevant laws and regulations; and efforts shall be made to generally permit low-risk AI in principle.

1.3. Current Status of AI Implementation in the Public Sector

Currently, there is no official data on what AI systems are being implemented to what extent by national institutions, local governments, public institutions, and public enterprises. The current status can only be estimated through research reports on public sector AI implementation.

1.3.1. Current Status of AI-Based Public Services

In October 2023, the National Assembly Research Service published a report titled <Analysis and Improvement Tasks of AI-Based Public Services>, which surveyed AI-based public services either in operation or planned across 17 metropolitan and provincial governments. The report revealed significant variations among local governments, from those operating diverse AI-based public services to others yet to implement any services.

The report categorizes AI-based public services into several types:

- Chatbots
- ChatGPT-based services
- AI-based public services for the elderly, disabled, and foreigners
- AI-based security monitoring
- River facility information systems and sewer pipeline defect detection systems

Chatbots emerge as the most widely implemented AI-based public service across local governments. Notable examples include Daegu's "Ddubot" (launched in 2017), Seoul's "120 Consultation Chatbot" (2019), Gyeongbuk's generative AI-based "Chat-Gyeongbuk" (2023), and Busan's welfare-focused chatbot "Self-reliance Honey Pot." Chat-Gyeongbuk, a ChatGPT-based service, supports various administrative tasks such as drafting press releases, recommending policy documents, analyzing government budgets, and writing project proposals. Services targeting the elderly and disabled include Gyeongnam's AI speakers, Jeju City's AI-IoT-based elderly health care program and intelligent civil service form assistance, and Daejeon's Smart Mirror (a civil service guidance system for the visually and hearing impaired). AI-based security monitoring enables CCTV control center operators to selectively monitor only footage showing movement. Daejeon City is implementing a "River Facility Information System," while Seoul is developing an AI-based "Sewer Pipeline Defect Detection System."

Among all local governments, the Seoul Metropolitan Government leads in implementing the most diverse range of AI-powered public services. Their comprehensive AI ecosystem extends beyond the 120 Consultation Chatbot to include an chatbot for employee work, an AI-powered meeting minutes system utilizing natural language processing technology, an intelligent video collaboration platform, rule-based workflow automation, and an advanced security monitoring platform.

1.3.2. Implementation Status of AI in the Public Sector

In 2022, the Software Policy & Research Institute (SPRI) conducted a survey of 408 public institutions (41 central administrative agencies, 17 local governments, and 350 public organizations) to examine their AI implementation status. Through surveys of staff members at each institution, they analyzed various aspects including AI application areas, purposes, technologies used, and barriers to AI adoption.

In 2024, SPRI published a new report titled <Study on AI Implementation Status in the Public Sector>. This time, they analyzed bidding and contract information from the Public Procurement Service's "KONEPS" platform over the previous decade (2013-2022), aiming to obtain more accurate and objective data compared to their previous survey-based research.

According to SPRI's findings, there were 3,870 AI implementation contracts during the ten-year period leading up to 2022. Of the 420 public institutions studied, 238 (56.7%) had implemented AI systems. The adoption of AI in public institutions has been rapidly increasing each year since the 2016 AlphaGo event. The number of AI-related contracts grew nearly sevenfold from 107 in 2015 to 922 in 2022, while the contract value increased from 93.8 billion won in 2016 to 1.08 trillion won in 2022. Before 2016, AI was primarily implemented for public services. However, the focus then shifted toward internal capacity building and operational efficiency. Since 2020, with advancements in technologies like chatbots and natural language processing, there has been a renewed emphasis on public services.

Regarding AI implementation areas, general administration - which includes e-government and civil service systems - has consistently maintained the highest share at over 20%. This is followed by industry/employment (industry development, job matching, etc., 16.5%), transportation/construction (intelligent transport networks, etc., 11.6%), and weather/disaster safety (weather prediction, etc., 10.4%). From a technical perspective, while OCR technology for document digitization and TTS (Text-to-Speech) technology for improving accessibility were predominantly used in the past, the application of other technologies has expanded since 2016. Chatbot implementations surged from 3 cases in 2016 to 161 in 2022, and with advances in voice recognition and unstructured data processing, the application of STT (Speech-to-text) and natural language processing technologies has also rapidly increased.

In terms of contract numbers, national institutions led (36.8%), followed by local governments (23.5%), quasi-government organizations (19.4%), and other public institutions (19.4%). However, the total contract value distribution showed a different pattern: national institutions (56.2%), quasi-government organizations (27.4%), local governments (11.7%), and other public institutions (11%), indicating relatively smaller contract sizes for local governments. National institutions primarily implemented AI in general administration, while also adopting AI solutions specific to their unique functions. Local governments similarly focused on general administration, followed by

transportation/construction for addressing local traffic issues, and increasing implementations in weather/disaster safety. Other public institutions, excluding local governments, showed twice as many AI implementations for internal capacity building compared to public services.

While this survey didn't classify AI systems by risk level, such a classification could have been valuable despite the lack of established social standards for AI risk assessment. This would help identify AI systems in the public sector that pose significant risks to safety and human rights. For instance, though detailed project specifics weren't examined, the following AI projects are expected to have significant potential impacts on safety and human rights:

• Gyeongsangnam-do's AI-based intelligent 119 emergency call reception system
• National Police Agency's enhancement of sexual violence victim investigation and support system using AI voice recognition
• Supreme Prosecutors' Office's development of big data-based intelligent digital evidence analysis platform
• Jeju's large bus driver drowsiness detection and response service
• Social Security Information Service's machine learning/RPA-based social service voucher fraud detection system
• Incheon Metropolitan City's data-based nighttime alley safety system
• Korea Customs Service's big data analysis model (BigFINDER) development

1.3.3. Plan for Promoting AI Implementation and Utilization in the Public Sector

On April 17, 2024, the Presidential Committee on Digital Platform Government held a general meeting to announce and discuss six policy initiatives, one of which focused on promoting AI implementation and utilization in the public sector. The government presented the following detailed strategies:
First, creating and spreading successful AI use cases in the public sector. This includes expanding support for advanced AI applications and providing focused support for large-scale projects across various areas, particularly in administrative efficiency and problem-solving.
Second, strengthening public sector AI capabilities. This involves distributing 'Guidelines for Implementing and Using Advanced AI in the Public Sector' and providing tailored training programs for practitioners using AI.
Third, establishing a government-exclusive advanced AI infrastructure. This includes developing a mid-to-long-term roadmap through Information Strategy Planning (ISP), along with preliminary preparations such as selecting pilot application targets and government training data.

On July 15, 2024, the Digital Platform Government Committee and the Ministry of Science and ICT announced eight selected projects for advanced AI public service development:

• Integrated R&D support service based on advanced AI
• Smart fire safety service
• Generative AI-based defense facility construction administration support

• AI labor inspector support service
• Specialized service for young farmers
• Advanced AI-based patent examination support service
• Multimodal advanced AI communication support service for people with disabilities
• Early detection support service for slow learners using advanced AI

1.3.4. AI Governance

The jurisdiction over public sector AI policy currently remains somewhat ambiguous. While the e-Government Act falls under the Ministry of the Interior and Safety, the Framework Act on Intelligence Informatization is overseen by the Ministry of Science and ICT (MSIT). The MSIT leads national-level AI policy initiatives and oversees the AI Basic Act, which passed the National Assembly in December 2024. However, neither the MSIT nor the Ministry of the Interior and Safety has conducted comprehensive surveys on AI implementation in the public sector.

Although the MSIT serves as the primary ministry responsible for AI technology, industry, and policy, the Yoon administration established the National Artificial Intelligence Committee under the President's office on September 26, 2024, as the highest-level governance body for deliberating and deciding major AI policies. This committee was launched through a presidential decree even before the enactment of the AI Basic Act. The committee is co-chaired by the President and a civilian representative, and includes 30 private sector members alongside government officials. However, while it includes representatives from industry, academia, and the legal profession, it notably lacks representation from civil society organizations.

On November 27, 2024, Korea launched its 'AI Safety Institute.' The institute was established to provide systematic and professional responses to various AI risks arising from technical limitations, human misuse of AI technology, and potential loss of AI control. This initiative aligns with the global trend of establishing AI safety institutes in major countries such as the UK and US following the AI Safety Summit held in the UK in late 2023.

1.4. Key Issues in Public Sector AI Implementation and Usage

1.4.1. Need for Public Sector AI Registration System

Currently in South Korea, since public institutions are not required to report or register AI system implementations, we can only track adoption status indirectly. To enable objective monitoring of AI implementations in public institutions at any time, consideration should be given to establishing an AI registration system. This would require public institutions to register key information when implementing AI systems, including their purpose, AI technologies used, functions, and development companies.

While consumers can choose not to use AI systems implemented by private companies, citizens cannot opt out of public services, making it difficult to avoid the impact of public sector AI systems. Even when public institutions implement AI for internal

operations, citizens may be unknowingly affected by AI-influenced policy decisions that impact their rights and obligations. Therefore, to identify and address potential risks of AI used in the public sector, we need information about which institutions are using AI, for what purposes, and with what capabilities.

Furthermore, many public sector organizations likely use AI systems for similar purposes. Sharing implementation status and understanding demands could prevent duplicate investments and budget waste. The Software Policy & Research Institute's report highlights this issue, noting that "many local governments are implementing AI systems sporadically without systematic planning and with limited budgets." The report suggests that "a strategy is needed where higher-level institutions take the lead in solving common regional problems using AI and then systematically disseminate solutions to lower-level institutions." In this regard, Article 7, Paragraph 6 of the Framework Act on Intelligence Informatization enables the Minister of Science and ICT to "establish measures to prevent duplicate investments in intelligence information projects pursued by national institutions."

1.4.2. Lack of Guidelines for Public Sector AI Implementation and Utilization

According to the SPRI's research, a significant number of AI systems - 3,870 procurement contracts over the decade leading up to 2022 - have already been implemented in the public sector, with more implementations expected in the future. However, it's concerning that there are no established guidelines specifying the principles and procedures that public institutions must follow when implementing AI systems.

While the Ministry of the Interior and Safety released a draft "Practitioner's Guide for AI Implementation in the Public Sector" for public consultation in April 2021, this has not yet been adopted as official guidance, and it's unclear to what extent public institutions are actually using these draft guidelines. The National Assembly Research Service has suggested in its report that "central government-level guidelines, planning, and coordination are necessary" as one of the future improvement tasks.

The 2024 "Guidelines for Implementing and Using Advanced AI in the Public Sector" represents a step forward. However, it focuses specifically on advanced AI-based systems and remains at a general level. There is a need for more detailed guidelines that cover all types of AI systems, including advanced AI, and incorporate specific evaluation criteria and procedures for assessing AI risks.

In comparison, the UK released its "Guidelines for AI Procurement" in June 2020. These guidelines present ten key considerations for public institutions when procuring AI systems, along with stage-by-stage procurement considerations. Korea similarly needs specialized procurement guidelines for AI systems.

1.4.3. Need for AI Human Rights Impact Assessment

While the <Draft Practitioner's Guide for AI Implementation in the Public Sector> requires consideration of AI ethics during system implementation, this document is not an official guideline, and the specific application of AI ethics is left to individual discretion. Similarly, the <Guidelines for Implementing and Using Advanced AI in the Public Sector> lacks concrete procedures for ensuring ethical compliance and identifying and mitigating risks.

When implementing AI systems in the public sector, human rights impact assessments should be conducted to ensure safety measures proportionate to the risks involved. The National Human Rights Commission of Korea recommended the introduction of 'AI Human Rights Impact Assessment' to the government and National Assembly in its 2022 'AI Human Rights Guidelines,' and released an 'AI Human Rights Impact Assessment Tool' on May 23, 2024, to facilitate voluntary assessments.

The U.S. Office of Management and Budget's (OMB) implementation guidance for President Biden's AI Executive Order of October 30, 2023, provides federal agencies with guidelines for AI implementation, including AI impact assessments, real-world testing, independent evaluations, and stakeholder consultations.In Korea, public institutions should conduct human rights impact assessments when implementing AI systems to evaluate risks and establish safety measures, and these procedures should be incorporated into official guidelines. To ensure thorough impact assessments, it's essential to guarantee the participation of people affected by the AI system, or civil society organizations and experts who can represent their interests.

Korean public institutions are required to establish and operate human rights management systems, including conducting human rights impact assessments. This framework should be extended to include mandatory assessments for AI system implementations. The AI Basic Act, passed by the National Assembly in December 2024, also imposes an 'obligation to strive' to conduct human rights impact assessments on high-impact AI operators.

1.4.4. Need for Legal Basis

When AI system implementation goes beyond merely improving existing work efficiency and significantly affects people's rights and responsibilities, additional legal basis may need to be established. For example, even if people's identities have been verified in immigration processes before, if biometric information is newly collected for identity verification or if AI systems are used for profiling to detect illegal immigration, these contents need to be additionally specified in relevant laws. This will justify the use of AI systems from a rule of law perspective and enable questioning of the appropriateness of implementation.

1.4.5. Recruitment and Education of Public Sector AI Experts

The National Assembly Research Service report suggests the need to raise awareness of AI

technology among public officials and citizens, and recommends education to improve public officials' data literacy and strengthen AI technology personnel. The SPRI report also suggests the need to cultivate internal AI experts as experts with both domain knowledge and AI knowledge are needed.

Chapter 2

AI in Law Enforcement: Implementation Status and Key Concerns

2.1. Introduction

Law enforcement powers must be exercised with limitations to protect public safety. AI systems in law enforcement particularly require social oversight, as they can serve state-of-the-art surveillance and authoritarian practices.

In South Korea, AI systems used by police and immigration authorities are being developed and deployed for public interest purposes such as crime prevention, criminal investigation, and immigration control. However, citizens who will be significantly impacted by these AI systems have no opportunity to provide input or have their concerns addressed. This is because decisions about the development and deployment of these AI systems are largely made behind closed doors. Citizens, as data subjects, are neither given the chance to consent nor even notified when their personal information is collected and used for the development of AI systems. It remains unclear whether proper legal foundations exist for deploying these AI systems or whether the Personal Information Protection Act is being strictly followed during their development and deployment. There is no governance system in place to control or monitor potential abuses of power by police and immigration authorities. It's also uncertain whether explainability and remedies will be guaranteed when these AI systems cause harm.

The police and Ministry of Justice appear focused solely on maximizing AI capabilities based on their institutional needs and industry demands, while neglecting human rights concerns. The situation is exacerbated by insufficient legal controls over the development and deployment of law enforcement AI. However, the development and deployment of AI systems for police and immigration purposes must be constitutionally controlled and balanced with human rights throughout their entire lifecycle. Specifically, AI systems that severely impact human rights should be prohibited, and even those that pose high risks should be subject to strict legal controls and regulations.

2.2. Law enforcement AI

2.2.1. Introduction

Korean law enforcement has been actively implementing AI technologies in public security operations, investing substantial budgets and fostering the development of related industries.

In November 2017, the "Fourth Industrial Revolution Response Plan," jointly announced by the National

Police Agency and the Ministry of Science and ICT, included the following initiatives for integrating intelligent technologies with public security infrastructure:
First, the development and demonstration of intelligent CCTV systems for identifying missing children and suspects (hereafter 'Police Intelligent CCTV Project'), 3D facial recognition, AI-based crime analysis, online obscenity blocking, and drone-based autonomous patrol and tracking systems.
Second, by 2020, the implementation of AI technology to analyze crime locations, types, and footage or identification data for suspects. The Police Intelligent CCTV Project is specifically referred to as "AI-based Complex Cognitive Technology."

Since 2019, the police began establishing a separate "Comprehensive Plan for the Advancement of Science and Technology in Public Security" (hereafter 'Public Security Technology Plan') on a five-year basis. Within the First Public Security Technology Plan (2019-2023), the initiative for "Development of Crime Prediction and Response Technologies Using Advanced Technology" overlaps with the public security technology tasks outlined in the previously mentioned Fourth Industrial Revolution Response Plan.

This initiative consists of three components using big data and AI:
1. Crime analysis and recognition technology
2. Crime prediction and prevention technology
3. Civil complaint response technology

Notably, the crime prediction and prevention technology component aims to "predict crimes through cameras including CCTV, track criminals through facial recognition, and predict dangerous behavior."

The Police Intelligent CCTV Project, developed under these initiatives and completed with testing in 2023, incorporates complex cognitive functions for identification, tracking, and predictive analysis of crimes and behaviors. However, the project overlooked a crucial consideration: international norms generally prohibit law enforcement agencies from using real-time biometric surveillance systems (such as facial or behavioral recognition) for remote identification and tracking in public spaces, classifying such technologies as extremely high-risk AI applications.

The Second Public Security Technology Plan (2024-2028) was developed after global foundation models like GPT-3.5 made significant worldwide impact. This plan includes a more systematic roadmap for developing and managing security technologies compared to its predecessor, going beyond merely addressing current issues. The police aim to enhance and nationwide expand their existing Intelligent CCTV project. Furthermore, they announced plans to build a real-time crime prediction and response system powered by AI, which would integrate and utilize a wide range of data for AI training - not only from police and related government agencies but also external data held by private entities.

In the sections that follow, we will examine the police's comprehensive strategy for implementing AI-based technologies, primarily focusing on the Second Public Security Technology Plan, and evaluate its implications for human rights.

2.2.2. Related Regulations

Police science and technology projects are based on the "Act on the Organization and Operation of National Police and Autonomous Police" (Article 33) and its Enforcement Decree "Regulations on the Promotion of Science and Technology in Public Security" (Article 3). However, these regulations serve as the legal basis for 'research and development' projects only.

Act on the Organization and Operation of National Police and Autonomous Police

Article 33 (Support for Research and Development Necessary for Public Security) (1) The Commissioner General of the Korean National Police Agency shall prepare and implement policies to promote science and technology in the field of public security, including the research, experiment, examination and development of technology required for public security (hereinafter referred to as "research and development projects") and training of specialists.

(2) To efficiently promote research and development projects, the Commissioner of the Korean National Police Agency may have any of the following institutions, organizations, etc. conduct research and development projects by entering into an agreement with them:

1. A national or public research institute;
2. Specific research institutes under Article 2 of the Specific Research Institutes Support Act;
3. Government-funded science and technology research institutes established under the Act on the Establishment, Operation and Fostering of Government-Funded Science and Technology Research Institutes;
4. A university, industrial university, junior college, or technical college under the Higher Education Act;
5. A public security research institute established as a corporation pursuant to the Civil Act or any other statutes, or a research institute annexed to corporations;
6. Research institutes annexed to enterprises or research and development divisions of enterprises recognized under Article 14-2 (1) of the Basic Research Promotion and Technology Development Support Act;
7. Other institutions or organizations prescribed by Presidential Decree, which conduct the research, examination, technology development, etc. related to public security.

(3) The Commissioner General of the Korean National Police Agency may fully or partially contribute or subsidize the expenses required by the institutions, organizations, etc. referred to in each subparagraph of paragraph (2) in implementing research and development projects.

(4) Matters necessary for the implementation of research and development projects under paragraph (2) and the payment, use, management, etc. of contributions referred to in paragraph (3) shall be prescribed by Presidential Decree.

Regulations on the Promotion of Science and Technology in Public Security

Article 3 (Establishment of Comprehensive Plan and Implementation Plan for the Promotion of Science and Technology in Public Security) (1) The

Commissioner General of the National Police Agency shall establish a comprehensive plan for the promotion of science and technology in public security (hereinafter referred to as the "Comprehensive Plan" in this Article) every five years as part of the measures for promoting science and technology in public security pursuant to Article 33(1) of the Act on the Organization and Operation of National Police and Municipal Police (hereinafter referred to as the "Act"). <Amended December 31, 2020>

(2) The Comprehensive Plan shall include the following matters:

1. Current status and prospects of science and technology in public security
2. Development direction and objectives of science and technology in public security
3. Analysis of domestic and international environment and measures to strengthen competitiveness of science and technology in public security
4. Strategic development of core technologies in public security
5. Training plan for professional personnel in science and technology in public security
6. Mid to long-term investment plan for the promotion of science and technology in public security
7. Other matters deemed necessary by the Commissioner General for the promotion of science and technology in public security

(3) The Commissioner General shall establish and implement annual implementation plans (hereinafter referred to as the "Implementation Plan" in this Article) in accordance with the Comprehensive Plan.

(4) The Implementation Plan shall include the following matters:

1. Implementation direction for the development of science and technology in public security for the relevant year
2. Detailed plans by sector for the promotion of science and technology in public security
3. Investment plans for major research and development projects in public security science and technology
4. Other matters deemed necessary by the Commissioner General for the promotion of science and technology in public security

(5) Matters necessary for the establishment and implementation of the Comprehensive Plan and Implementation Plan other than those prescribed in Paragraphs (1) through (4) shall be determined by the Commissioner General.

In order to deploy the technical achievements of research and development in actual law enforcement settings, proper legal grounds must be established under relevant laws governing police activities and the Personal Information Protection Act. However, it remains unclear whether there are adequate legal grounds in the process where the police, after completing the research and development of advanced technology prototypes and conducting demonstration tests and pilot installations, proceed to actual deployment.

2.2.3. Current Status

The police have been developing and deploying various advanced AI technologies, and through the Second Public Security Science and Technology Plan, effective from 2024, they aim to establish technical, material, institutional, and governance frameworks for more extensive AI technology deployment.

First, in terms of "crime prevention and response systems," the police plan to develop a 'real-time' crime response system. To address the recent increase in random motivation crimes (indiscriminate violence), they are developing models to 'analyze and predict crime patterns' by analyzing historical crime data. Additionally, for 'proactive crime prediction, real-time identification, and response,' they are developing algorithms to analyze behavioral patterns of stalkers and sex offenders and to assess 'individual risk levels.' They also detect abnormal risks by recognizing 'behavior' and 'voice' in real-time in public places through patrol cars, robots, drones, and CCTV. The police aim to design comprehensive 'intelligent risk indicator analysis systems' and 'intelligent integrated control systems' by linking these criminal behavior pattern analyses and abnormal behavior detection technologies.

Ultimately, the police aim to establish a Korean-style RTCC (Real Time Crime Center) that will command or support linked or integrated data from various agencies and real-time crime prediction capabilities. For intelligent real-time integrated control, they are linking and integrating various security data held by the police with data from various relevant agencies. Target external agency data includes Open Source Intelligence (OSINT), Geographic Information Systems (GIS), CCTV, and assembly and crowd concentration risk simulations based on digital twin convergence of national spatial data. This integrated data is analyzed in 'real-time' and used to develop technologies for predicting crime occurrence possibilities and patterns.

Meanwhile, the police plan to develop multimodal multi-video multi-object recognition and matching technology that automatically identifies and tracks specific individuals, including socially vulnerable individuals and criminals, through movement pattern and object recognition in public places. This represents an advancement of the existing police intelligent CCTV project.

The police intelligent CCTV project, which remotely identifies and tracks individuals through real-time biometric recognition including facial and behavioral recognition in public places, completed development during the First Public Security Science and Technology Plan period and finished demonstration at the Anyang City Integrated Control Center in Gyeonggi Province in November 2023. This project consists of △Complex Cognition Core Source SW Technology (Ministry of Science and ICT), △Wearable Devices for Identity Verification (Ministry of Industry), and △Complex Cognition Technology Application and Infrastructure (National Police Agency), with a total budget of 32.5 billion won. From the 2024 business plan, an additional 750 million won is being invested to advance this technology and implement it nationwide by linking it with the National Police Agency system.

In terms of "police equipment modernization," the police plan to introduce drones. The scope of drone applications will expand from traffic management and enforcement to regular patrols and even documentation of assemblies and demonstrations. Additionally, they are developing unmanned patrol robots capable of autonomous outdoor patrol using quadrupedal locomotion and other methods. To facilitate this, they are pursuing the establishment of "Police Patrol Robot Operation Rules." For civil service responses, they plan to develop interactive

chatbots and face-to-face robots equipped with risk response capabilities.

Furthermore, the police plan to advance their "science and technology infrastructure." They aim to enhance their forensic science platform using AI to comprehensively analyze forensic data. This includes technology that automatically refines and standardizes data needed by the police, as well as technology for 'real-time' response to voice phishing conversations in cooperation with telecommunications companies.

The police are also planning countermeasures against "AI-utilized cybercrime." This includes deploying AI to detect phishing sites in 'real-time' and automatically detect fake news. They are also developing cyber patrol technology to monitor conversations and posts across various internet platforms in 'real-time.' To detect mobile financial fraud, they are developing technology to identify criminals' 'voices' and recognize victims' 'emotions.'

Additionally, the police have plans to amend the Intelligent Robots Act to implement mandatory safety certification and accident recording for robots as part of creating a "traffic safety system" that aligns with advanced mobility environments.

The Second Public Security Science and Technology Plan particularly emphasizes creating an innovative foundation for police and security industries through security science. To facilitate the police's digital transformation, they have planned to integrate information systems currently separated into over 100 task-specific systems and improve data utilization. The police are taking a more proactive approach to building a data lake than before, linking structured and unstructured data within the National Police Agency, purchasing external data, and receiving data through MOU partnerships. As of 2024, the police big data platform (data lake) has established a foundation for shared utilization by linking 297 types of data - 125 types of internal data dispersed across various departments of the National Police Agency and 172 types of external data. The police also plan to establish a new data center where collected data can be used for AI learning to promote the development of police AI models.

2.2.4. Issues

While establishing a scientific crime prevention system and conducting investigations based on accurate evidence could enhance public safety, there are significant concerns. If the police process excessive amounts of personal information and conduct surveillance beyond legitimate data processing purposes under the pretext of crime prevention and investigation, this could violate fundamental human rights protected by the Constitution. Furthermore, the police's use of AI surveillance and tracking technologies in public online and offline spaces without proper legal authority poses significant risks to human rights, as it could have a chilling effect on innocent citizens' freedoms. In this regard, the AI security technologies outlined in the Second Public Security Science and Technology Plan raise numerous concerns.

First, there is a severe lack of transparency to the public. While certain public security technology solutions are swiftly implemented after the establishment of the comprehensive plan, there are

virtually no procedures for citizens outside the police force to submit or have their opinions reflected in this comprehensive plan.
The comprehensive plan is reportedly established through the following process: formulation by the police's internal General Planning Committee, expert review, public surveys, public hearings, and deliberation by the National Science and Technology Advisory Council. The annual implementation plans are established through human rights impact assessments conducted by internal police organizations, reports to the Human Rights Commission, and resolutions by the National Police Commission. However, survey results are only cited as justification for the comprehensive plan without including any critical opinions. Information about public hearings is barely disclosed, either before or after they are held. No information is released about data processing or algorithm design directions. Consequently, there is effectively no channel through which ordinary citizens who will be affected by these projects, or human rights organizations representing them, can access information about these projects and submit their opinions.

Second, there are concerns about how the police handle personal data throughout the Public Security Science and Technology Plan. This is because the police have indicated their intention to not only maximize the integration of various security data held by police nationwide for vague purposes but also link it with external data held by other agencies or private entities and use it for AI training. For example, the police announced that they would establish "a system that integrates security data held by the police and various data held by relevant agencies" to build the Korean-style RTCC system. Furthermore, the police have plans to integrate systems and data that are currently separated by operational purposes and build an extensive data lake for undefined police purposes beyond the RTCC. The specific nature of the security data held by the police is not clearly known. They only state that it is "data that can be utilized under the law" and will be "used only for limited purposes such as public safety and crime prevention."

However, this data includes personal information, and the personal information held by the police encompasses not only criminal history records but also sensitive information such as thoughts and beliefs, labor union and political party membership and withdrawal, and political views. While the police have plans to extensively use and integrate such potentially sensitive personal information, they have not disclosed how they intend to protect this personal information. The comprehensive plan only briefly mentions establishing relevant legislation and conducting training regarding privacy concerns related to wearable equipment and drone operations. This raises doubts about whether the police's artificial intelligence projects have established and implemented plans to comply with the Personal Information Protection Act when collecting or using personal information. In particular, the legality assessment and supervisory mechanisms related to personal information collection, use, and processing appear to be very weak. There are concerns that the police might try to circumvent the Personal Information Protection Act by claiming pseudonymization, or deploy AI algorithms directly in police operations after completing extensive data training without proper constraints under the pretext of research and development.

On February 16, 2023, the German Federal Constitutional Court ruled that automated personal information analysis or evaluation by the police, including predictive policing, was unconstitutional due to violation of proportionality of legal interests (1 BvR 1547/19, 1 BvR 2634/20). The Court held that when personal information stored by the police is automatically processed for individual analysis or evaluation, this fundamentally restricts the right to informational self-determination for all individuals whose information is used in this process. While it is not uncommon for police to use previously obtained information as investigative leads, methods that process large amounts of complex information, such as automated analysis or evaluation, have a significant impact on fundamental rights. Therefore, for such additional processing of personal information by the police to be constitutionally justified, it requires additional grounds under the principle of purpose 'modification.' The predictive policing laws of Hesse and Hamburg states were ruled unconstitutional because they allowed unlimited processing of unlimited datasets with no restrictions other than the requirement "to prevent criminal acts." This means that even for personal information held by the police, using it for additional analysis or evaluation requires proving the existence of a specific risk. In particular, data from residential surveillance, online searches, communication interception, and traffic information inquiries should not be used for such data analysis.

However, Korean police plan to integrate potentially sensitive personal information on a large scale for vague police purposes simply because it is held by police agencies, use it for broad police purposes, and further integrate it with data from other agencies or private entities. This is being done without any specific legal basis or plans in place. This could violate constitutionally protected rights to personal information.

Third, the objectives of AI algorithms emphasizing 'real-time' detection and analysis, and 'predictive policing' are potentially violative of human rights. The police present 'real-time' detection and analysis as goals in various projects, which poses significant risks to human rights as it constitutes intensive surveillance. Real-time detection of location and conversations in both offline and online spaces also has a high possibility of violating communication privacy. Furthermore, the 'predictions' probabilistically generated by police AI could lead to unacceptable discrimination and human rights risks in our society. Analysis or evaluation of individuals based on biased group analysis could result in discriminatory outcomes for certain groups, such as specific races, and these results could be used again in learning, creating a 'feedback loop' problem.

In particular, police activities that remotely identify people in public places using biometric information such as facial features or movements in real-time are prohibited under international human rights norms. The EU AI Act, which took effect in August 2024, principally prohibited such practices because they "affect the privacy of many people, create a sense of being under constant surveillance, and could indirectly infringe on freedom of assembly and other fundamental rights." It only allows limited exceptions with prior court approval for extremely exceptional cases such as victim identification and counter-terrorism. Even before this, in 2021, the UN High Commissioner for Human Rights had called on governments to impose a moratorium on facial recognition lacking legal controls. Meanwhile, the

EU AI Act also prohibited predictive policing based solely on individual characteristics. It mandates that AI crime prediction must be based on objective and verifiable facts directly related to criminal activity.

Fourth, it is concerning that the Police Science and Technology Plan presents the promotion of the security industry as a major objective. The police announced that the First Police Science and Technology Plan achieved economic results of 428.5 million won through technology transfers, and secured export contracts worth $3.98 million (5.17 billion won) through industry exhibitions in 2023. The Second Police Science and Technology Plan contains more ambitious plans to contribute to the advancement of the security industry. The police are pursuing public-private partnerships between law enforcement, industry, academia, and research institutions, centered around a statutory agency specializing in police science and technology research (Korea Institute for Police Technology) and a private business association (Security Industry Promotion Association) supported by the police. Furthermore, they plan to enact the "Security Industry Promotion Act" to nationally support and manage not only domestic application but also overseas exports of security technologies.

However, police technology involves using citizen data held by police or public institutions directly for their own AI training purposes. Using such data to support private companies' development of security products beyond legitimate police purposes with legal grounds, could constitute an unjustified and disproportionate violation of personal information self-determination rights. Moreover, if the motivation and direction of security technology development become skewed toward commercial profit rather than public interest, it may produce technologies that emphasize invasive and excessive detection capabilities instead of protecting citizen safety through proportionate exercise of state authority. If the goal-setting for police AI, which significantly impacts citizens' human rights, takes place behind closed doors between police and industry without public oversight, it could pose a major threat to democracy. We should remember the criticism that the military-industrial complex during the Cold War era has threatened American democracy and the peace system to this day.

Fifth, there is currently a void in independent human rights oversight regarding the development and deployment of police AI systems that could negatively impact human rights. While certain police investigative techniques may require confidentiality, it is possible to establish professional and independent human rights oversight governance for police operations. The current human rights impact assessment conducted internally by the National Police Agency does not meet these conditions. Independent human rights oversight governance should be able to independently review the legality, data, and algorithms of police AI systems, as well as monitor compliance with the Constitution and Personal Information Protection Act. It goes without saying that this oversight should participate from the initial planning stages of police science and technology, providing checks and balances on police technology development and deployment, and remedying human rights violations. Above all, there needs to be legally established governance and control procedures in police operation laws to comprehensively oversee the development and deployment of police AI systems that significantly impact citizens' human rights.

2.3. Immigration AI

2.3.1. Introduction

Since April 2019, the Ministry of Justice has been developing an "Artificial Intelligence Identification and Tracking System (hereafter 'Immigration AI')" for deployment at airport immigration checkpoints. In this project, the Ministry of Justice provided multiple private companies with 170 million pieces of personal information collected from both Korean citizens and foreign nationals for AI training purposes, including facial data gathered for immigration control.

The project authorities - the Ministry of Science and ICT and the Ministry of Justice - as well as the Personal Information Protection Commission claim that this project did not violate the Personal Information Protection Act or any related laws and poses no issues. Nevertheless, the Ministry of Justice abruptly destroyed the training data and experimental lab, thereby denying requests from both Korean and foreign data subjects who sought to verify whether their personal information had been used. As of December 2024, this case is under constitutional review.

2.3.2. Relevant Regulations

The Ministry of Justice presented the following provisions of the Immigration Act as the legal basis for Immigration AI. The Personal Information Protection Commission endorsed these provisions as constituting a valid legal basis for Immigration AI.

Immigration Act

Article 1 (Purpose) The purpose of this Act is to provide for matters concerning safe border controls through the immigration control of all nationals and aliens who enter or depart from the Republic of Korea, control over the stay of aliens in the Republic of Korea, and social integration, etc.

Article 3 (Nationals' Departure from the Republic of Korea)
(5) Immigration control officials may utilize the biometrics information collected or submitted under paragraph (3) for departure inspections.

Article 6 (Nationals' Entry into the Republic of Korea)
(6) Immigration control officials may utilize the biometrics information collected or submitted under paragraph (4) for entry inspections.

Article 12-2 (Provision of Biometrics Information at Time of Entry)
(5) An immigration control official may use the biometrics information provided or submitted under paragraph (1) or (3) for entry inspections.

Article 28 (Departure Inspections)
(6) Immigration control officials may utilize the biometrics information provided or submitted under Article 12-2 (1) or (3) for departure inspections.

However, the Immigration Act only stipulates that biometric information can be used "for departure screening" or "for entry screening." This merely means that biometric information can be matched on a one-to-one basis to verify the identity of travelers. Therefore, the Immigration Act can hardly serve as a legitimate and legal basis for using personal information of Korean citizens and foreign travelers,

including facial data, to develop an AI system that automatically identifies and tracks individuals on a one-to-many basis.

2.3.3. Current Status

The Immigration AI is a joint project between the Ministry of Justice and the Ministry of Science and ICT, which outlined the following main objectives. First, to "advance the existing fingerprint-based airport immigration control system to a data and AI-based system." Second, to "contribute to the early acquisition of computer vision technology and stimulate the domestic AI industry by providing AI companies with public sector verification and market demand." Under these objectives, the government provided immigration data to multiple applicant companies for training purposes, and private companies trained their AI algorithms in a government-established experimental lab.

From the outset, this project's primary goal was to enhance the domestic and international competitiveness of private companies' facial recognition AI by using training materials such as facial image data provided by the Ministry of Justice. The National IT Industry Promotion Agency (NIPA), a public institution under the Ministry of Science and ICT that oversees the Immigration AI project, stated that "the facial data held by the Ministry of Justice is valued between 500 billion to 1 trillion won." They claimed this project could solve the problem where "collecting training facial data requires individual consent and incurs collection costs of 20,000 to 100,000 won per person, plus processing and management burdens." NIPA anticipated that this project would provide a foundation for AI companies' growth by "addressing their challenges" through "securing large-scale public data." Furthermore, they projected that the facial recognition technology developed through this project "would expand into various business areas including banking transactions, shopping, finding missing children, and identifying illegal immigrants."

From April 2019 until just before the project's issues were publicized and halted in October 2021, personal information stored for immigration purposes was provided for this project. A total of 12 applicant companies used personal information consisting of 57.6 million Korean citizens and 120 million foreign nationals for training their facial recognition AI. This data, amounting to approximately 170 million records, was selected through 'filtering' the original 320 million immigration records based on facial photo size, capacity, and file integrity. The personal information provided to companies included not only facial image data but also passport numbers, nationality, birth year, and gender. The provided facial image information was 'preprocessed' to be converted into machine-readable facial recognition data and classified by continent of origin and age groups. The applicant companies had not even established clear service contracts with the government. The government provided AI training data to multiple companies without contracts, citing the reason that they would later select companies with superior performance.

Meanwhile, in addition to the above immigration data, the Ministry of Justice had installed hundreds of motion-detecting CCTV cameras at airports and was separately collecting 'real data' from actual airport users. This was for developing an 'abnormal

behavior' detection system. However, as soon as the Immigration AI project became public knowledge, the Ministry of Justice removed the CCTV cameras collecting real data in airports. The Personal Information Protection Commission determined that while the installation and operation of CCTV itself could be justified for crime prevention and facility safety purposes, separate legal grounds were required for biometric recognition such as facial and motion detection. However, they did not make a determination on its legality, citing that the CCTV footage was never used as the project had been suspended.

After this issue was brought to public attention through media coverage and parliamentary audit in October 2021, civil society groups took action to criticize the human rights violations of Immigration AI and seek remedies for those affected.

However, the Personal Information Protection Commission, which investigated and reviewed whether this project violated the Personal Information Protection Act, decided that Immigration AI did not violate the Act as it was based on the Immigration Act. They merely imposed a minimal fine of 1 million won on the Ministry of Justice for delaying the notification of personal information processing outsourcing.

With the help of civil society organizations, about 20 Korean and foreign data subjects requested access to verify whether their personal information had been used as training data for this project. However, the Ministry of Justice rejected these requests, claiming that individuals could not be 'identified' by facial data alone. In response, the complainants filed for dispute mediation with the Personal Information Dispute Mediation Committee regarding the denial of their right to access. However, this request was dismissed as the Ministry of Justice had destroyed all data used for AI training and closed the experimental lab.

Subsequently, civil society organizations filed a public interest audit request with the Board of Audit and Inspection regarding the illegality of this project. However, the Board also closed the case, citing insufficient grounds for the audit request. Finally, in July 2022, civil society organizations filed a constitutional complaint with the Constitutional Court. As of December 2024, this case is under constitutional review.

2.3.4. Issues

The primary issue with the immigration AI project is its ambiguous legal foundation. While the Ministry of Justice and Personal Information Protection Commission claim that the Immigration Act allows the government to provide extensive personal data to companies for AI training and development, the Act only stipulates that biometric data can be processed for one-to-one identity verification during immigration procedures. This legal basis alone appears insufficient to justify using immigration data for other purposes, specifically developing AI systems capable of automatically tracking subjects in motion and identifying faces and behaviors.

Secondly, despite using vast amounts of personal data for AI training, the immigration AI project proceeded without obtaining consent or even providing basic notification to data subjects.

Sufficient measures were not taken to protect data subjects or guarantee their rights. Individuals had no opportunity to object to their personal data being used for training.

Facial recognition data, in particular, qualifies as sensitive biometric information requiring enhanced protection. Nevertheless, the Ministry of Justice cited irrelevant provisions of the Immigration Act to justify providing companies with sensitive data on a massive scale. Meanwhile, they denied data subjects' access requests, claiming individuals couldn't be identified by facial images alone. They later destroyed the data entirely, preventing affected individuals from verifying damages or pursuing civil litigation for remedies. This demonstrates a serious lack of accountability from a government agency implementing a large-scale national project based on extensive personal information of both domestic and international individuals.

Thirdly, the immigration AI's goal and methods of identifying and tracking people through facial recognition in public spaces is a human rights violation. AI capabilities that automatically identify and track faces represent unprecedented technology that could pose serious risks to human rights. The EU AI Act principally prohibits law enforcement agencies from remotely identifying individuals in public spaces using real-time biometric data like facial features or movements.

Fourthly, the project was fundamentally flawed in its purpose and method of providing vast amounts of government-held immigration data to multiple domestic facial recognition companies. The intended use and contractual relationships for algorithms trained by participating companies remain unclear. This validates criticism that the project prioritized industry promotion through data sharing over immigration purposes.

The immigration AI was a national project developing high-risk AI for real-time remote facial recognition in public airport immigration areas. Yet its development process was highly opaque and violated the rights of both data subjects whose personal information was used for AI training and domestic and international citizens who would be affected by the system. In conclusion, the immigration AI development project was driven solely by government and industry demands, posed significant human rights risks, and lacked accountability.

Chapter 3

Current State of AI in Education

3.1. Introduction

There has not yet been a comprehensive survey on how and for what purposes AI is being utilized in education. In the private education sector, AI-enabled learning applications (such as English learning apps and practice test assistance apps) are being developed and purchased by individuals for learning purposes. Teachers may individually use generative AI or AI services for creating test questions and lesson preparation. Some schools incorporate publicly available AI services into their classes, and certain universities are experimentally developing and implementing AI lecture systems.

As of December 2024, the Ministry of Education's only officially implemented AI system is 'AI PengTalk', developed for elementary students' English speaking practice. However, when the Ministry announced plans in 2023 to introduce AI digital textbooks(hereinafter AIDT) by 2025, it faced significant opposition from educators and parents. On December 26, 2024, opposition party-led legislation passed through the National Assembly, legally classifying AIDT as educational materials rather than official textbooks and allowing schools to choose whether to use them. The Ministry of Education is strongly opposing this development, threatening to request a presidential veto. Even if a veto is exercised and the bill is effectively rejected, the confirmed strong public opposition suggests that implementing the AIDT initiative will face significant challenges.

3.2. AI Norms and Regulation in Education

3.2.1. AI Ethics Principles in Education

On August 11, 2022, the Ministry of Education announced the "AI Ethics Principles for Education." This followed the government-wide "AI Ethics Standards" released in December 2020 and the National Human Rights Commission's "Human Rights Guidelines for AI Development and Implementation" published in May 2022. The Ministry explained that existing AI ethics standards were mostly generic and developer-centric, making them difficult to apply directly in educational settings. Therefore, they developed ethics principles specifically tailored for education.

The "AI Ethics Principles for Education" established under the main principle of "AI Supporting Human Growth" presents ten detailed principles:

1. Foster potential for human growth
2. Ensure learner autonomy and diversity
3. Respect educator expertise

4. Maintain strong relationships between educational stakeholders
5. Guarantee equal educational opportunities and fairness
6. Strengthen educational community solidarity and cooperation
7. Contribute to social public good
8. Ensure educational stakeholders' safety
9. Guarantee transparency in data processing and maintain explainability
10. Use data purposefully and protect privacy

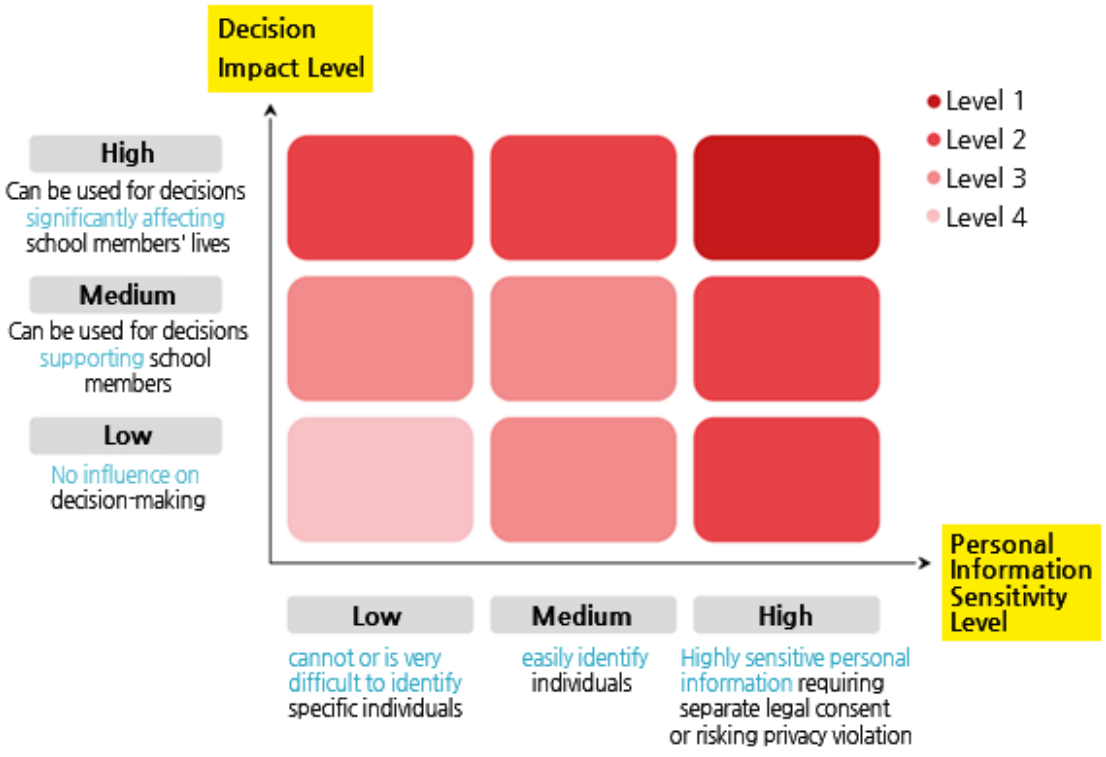
However, these principles remain abstract without providing specific procedures or guidelines for AI development and operation in education. Some explanations seem to soften the principles, likely considering AI industry concerns. For example, the safety principle notes that "care should be taken not to inadvertently sacrifice individuals or groups, while ensuring safety measures don't hinder technological advancement." Similarly, the transparency principle states that "while data processing transparency is a priority, efficiency considerations can be factored into achieving transparency."

It's questionable whether the AIDT initiative actually adheres to these principles. The government's insistence on pushing forward despite opposition from teachers, parents, and the opposition party contradicts the principle of "maintaining strong relationships between educational stakeholders." The rushed implementation without adequate preparation or pilot testing conflicts with the principle of "ensuring educational stakeholders' safety."

3.2.2. Field Guidelines for Securing AI Public Interest

In August 2021, the Seoul Metropolitan Office of Education released "Field Guidelines for Securing AI Public Interest," preceding the Ministry's ethics principles. These guidelines require schools to evaluate AI systems' risk levels and conduct detailed impact assessments before implementing web-based machine learning tools, AI tutors, AI speakers, chatbots, or facility management AI systems, etc.

[figure 1] **AI Risk Assessment Matrix**



AI systems are classified into four levels based on their decision-making impact and personal data sensitivity, with higher-risk systems requiring more rigorous procedures and review. The guidelines provide specific criteria for level assessment and a checklist for conducting AI impact assessments.

These guidelines are more practical than the "AI Ethics Principles for Education," offering concrete implementation guidance and an impact assessment checklist, albeit at a basic level. However, research is needed to evaluate their actual adoption in educational settings.

3.3. AI Digital Textbook Overview

According to the Ministry of Education, AIDT are "textbooks equipped with various learning materials and support functions using intelligent information technology, including artificial intelligence, to support personalized learning opportunities tailored to individual students' abilities and levels." The Yoon Suk-yeol administration's Ministry of Education pushed for their implementation. However, education experts expressed concern about AIDT, noting that its 'knowledge injection through repetitive learning' approach would inevitably lead to rote teaching methods, as teachers would be forced to follow the textbook's instructional style. The initiative faced opposition from the Korean Teachers and Education Workers Union, parent associations, civil society organizations, and opposition parties.

3.3.1. Progress of AIDT Initiative

In February 2023, the Ministry of Education announced its "Digital-Based Educational Innovation Plan" with the slogan "Realizing Customized Education for All." The plan aimed to create an "optimized customized education system based on individual students' capabilities, preferences, and learning pace." The Ministry asserted that "AI and advanced technologies can improve educational quality" and that "the digital transformation era demands fundamental changes in educational content and methods." AIDT were proposed as a key tool for this educational innovation.

On June 8, 2023, the Ministry of Education announced its "AI Digital Textbook Implementation Plan." The rollout would begin in 2025 with mathematics, English, information technology, and Korean language (special education) for grades 3-4 elementary, 1st-year middle school, and 1st-year high school students. The plan aimed to expand gradually to all subjects including Korean language, social studies, and science by 2028, excluding grades 1-2 elementary, high school electives, arts/physical education, and moral education due to developmental considerations.

Under this plan, the government and public institutions would build an integrated learning repository (including unified login and dashboard), while private companies would develop subject-specific digital textbooks. The policy expanded textbook development rights beyond traditional publishers to include EdTech companies forming consortiums with publishing houses. These EdTech companies would be eligible to apply independently after 2029.

On October 17, 2023, the Ministry of Education amended the "Regulations on Curriculum Books" to

establish the legal status of AIDT. The revised enforcement ordinance included definitions and certification requirements for digital textbooks, defining them as learning support software using intelligent information technology (Article 2, Clause 2). The certification process would also evaluate technical defects and service compatibility.

The certification review for AIDT, initially planned for early 2024, was delayed until September 2024. The main review results were announced on September 24, followed by an appeals process and revision review, with final approval results announced on November 29. In total, 76 textbooks from 12 publishers passed the certification process.

3.3.2. Key Contents of the AI Digital Textbook Development Guidelines

On August 30, 2023, the Ministry of Education announced the "Artificial Intelligence (AI) Digital Textbook Development Guidelines" (hereafter referred to as "the Guidelines"). According to the Guidelines, a portal will be established to provide AIDT, and an integrated authentication system will be implemented to allow users to access both the AIDTportal and each publisher's digital textbooks with a single account.

The AIDT will provide comprehensive diagnostic analysis of individual students, including their strengths, weaknesses, and learning attitudes. Based on the analysis of students' learning comprehension and characteristics, the system will present personalized learning paths and content tailored to each individual's abilities and goals. Through the collection and analysis of learning data using AI technology, the system will offer a so-called 'AI tutor' service that supports individualized learning customized to each student's unique characteristics.

To provide personalized content to students, the system analyzes students' interests, proficiency levels, and learning situations, which may be based on sensitive personal information. According to the Guidelines, they provide an example where "for students who enjoy speaking English, voice recognition-based conversation simulation content is provided, recommending learning content that considers individual learning patterns to help students maintain continuous engagement in learning."

Looking at the student dashboard example presented in the Guidelines, it is anticipated to collect and analyze extensive personal information including:

- Basic personal information
- Learning engagement metrics (login/logout times, number of logins, number of content pieces studied, study time, number of posts)
- Academic achievement (assessment results, assignment submission status, number and results of problems solved, existence and number of incorrect answer notes)
- Learning history (recently studied units, recently attempted problems, solved problems, clicked content)
- Learning analysis (challenging units, learning map recommendations, learning competencies, learning patterns such as login intervals or study time)

The dashboard indicates that the AIDT not only monitors all aspects of a student's learning process

during class hours but also after school, and even includes features to detect students' 'moods'. Whether this mood detection is implemented through 'emotion recognition' technology needs to be examined by reviewing the actual functions of the AIDT. Through the parent dashboard, parents can access their children's data and receive information about their children's moods.

The 'AI tutor' provides various functions including question-and-answer support, additional learning material provision, learning strategy suggestions, learning progress monitoring, feedback and achievement assessment, and incorrect answer notes. It includes a feature that responds to students' questions when they have inquiries, which can be implemented in various forms depending on the developer, such as chatbot-style or voice recognition-based systems.

The 'AI Teaching Assistant' service provides teachers with information about each student's learning activities and supports customized lesson planning. Teachers can establish individual learning plans based on student data and provide personalized learning paths for each student based on learning activity monitoring results.

The AIDT is provided as a cloud-based (SaaS) web service, which requires the use of infrastructure (IaaS) and software (SaaS) with Cloud Security Assurance Program (CSAP) 'medium' grade certification or higher. The collection and storage of student learning data is carried out by the publishers. Publishers will utilize the learning data generated from the AIDT independently to provide subject-specific learning analytics information.

The Guidelines require the following elements to be included when establishing data management policies that involve students' personal information:
- Ensuring Reliability: Data collected through AIDT must be accurate and reliable. Since learning data containing errors or distorted information can degrade the performance of learning algorithms or lead to inaccurate results, data accuracy must be verified and, if necessary, undergo data cleaning processes. Additionally, data quality and risks must be managed to minimize data bias throughout the entire process of learning data collection and utilization.
- Consent for Data Usage: To use student learning data, explicit consent must be obtained from students (or parents) through consent forms specifying the time of data collection, purpose of data usage, scope, and duration.
- Data Security: Learning data generated during the use of AIDT must be managed by developers in compliance with strict security and safety regulations. This requires adherence to administrative and technical security guidelines.
- Prohibition of Use Beyond Intended Purpose: Data collected through AIDT must be used only for the purpose of improving AIDT services and must not be used for publishers' own services. Furthermore, infrastructure for proprietary services must be managed separately from AIDT infrastructure.
- Data Provision: Appropriate responses must be made to requests for data transfer from students (or parents) in accordance with data portability rights.
- Personal Information De-identification: Learning data for national-level learning analysis must undergo de-identification measures (anonymization, pseudonymization).
- Data Retention and Disposal: Developers must manage and retain student learning data collected from AIDT in accordance with the policies outlined in

the personal information usage consent form. When the data retention period expires, the retention period may be extended with user consent, or the data must be safely disposed of.

The Guidelines require the establishment and compliance with processes for AI risk management when developing AIDT:
- Safety and Inclusivity Management: AIDT must not include content that contains sexually explicit or violent material, or content that damages or distorts social values.
- Fairness Management: AI algorithms must not produce biased results through learning data.
- Algorithm Transparency: AIDT must be able to explain how personal information is used and how specific decisions or actions are performed.
- Resonsibility Management: Clear responsibility must be established for specific AI decisions or actions.

However, these points are merely guidelines that AIDT publishers need to consider. Further investigation is needed to understand how AIDT that have passed the Ministry of Education's certification process have technically implemented the functions outlined in the guidelines, and what procedures they have followed to address personal information protection and AI risk management principles.

3.3.3. Certification Process for AIDT

The certification process for AIDT is divided into two main parts: content review and technical review. Content review is conducted by subject-specific certification institutions - the Korea Foundation for Science and Creativity handles mathematics, while the Korea Institute for Curriculum and Evaluation is responsible for English and information technology. The technical review is carried out in cooperation with the Korea Education and Research Information Service (KERIS) as the technical review support agency, using the 'Self-Technical Verification Report' submitted by developers as baseline material. For the Self-Technical Verification Report, developers either commission external organizations or submit detailed documentation of their own usability checks and specification test results according to technical review criteria. AIDT that pass the certification review undergo field suitability testing with actual users including students and teachers, and may be ordered to make modifications if necessary. The technical review areas and criteria are broadly categorized into usability testing, technical standards compliance testing, compatibility testing, and reliability testing, with specific review items and elements as follows:

[table 1] **AI Digital Textbook Certification Technical Review Standards and Guidelines**

Review Area	Review Items	Review Elements	Content Review	Technical Verification
Usability (10)	• Does the AIDT operate without technical defects or errors?	• Digital textbook functionality error verification • AI performance testing • Load testing	O	-
	• Does the AIDT ensure web accessibility and interoperability?	• Compatibility between devices and browsers • Ease of access for students with disabilities • Multi-language support through automatic translation	O	-
Technical Standards Compliance (10)	• Does the technology included in the AI digital textbook comply with relevant specifications and standards?	• Meeting infrastructure environment requirements • Compliance with relevant standards	O	-
	• Does it reflect the compliance requirements for AIDT development?	• Compliance with measures prohibiting use beyond intended purposes • Compliance with measures prohibiting 'Learning ahead of the curriculum' • Compliance with and measures for AI ethics • Copyright security	O	-
Appropriate-ness Testing (40)	• Are the linking features of the AIDT appropriately configured?	• Authentication system • Application of curriculum standard framework • Initial screen configuration	O	O
	• Do the AI-based personalized learning support functions operate appropriately?	• Learning diagnosis and recommendations • Dashboard and data visualization • AI tutor for learning guidance and support	O	O

		• AI teaching assistant for lesson design and instruction • Support for teacher content restructuring functions		
	• Is the UI/UX design and interaction of the AI digital textbook configured conveniently from a user perspective?	• UI/UX usability • Appropriateness of interactions	O	O
Reliability (40)	• Is data appropriately collected and securely managed?	• Data collection & storage • Data transmission	O	O
	• Is the personal information and information security system operated reliably?	• Personal information protection, prevention, and response measures • Information security, prevention, and response measures • Operation of related physical/technical systems	O	O
	• Is user support and service management operated reliably and stably?	• Operation of user support and response management system • Operation of error management system • Configuration of service quality management system	O	O

3.4. Problems with AIDT

3.4.1. Insufficient Stakeholder Consultation and Preparation

One of the major issues with AIDT is that the Ministry of Education is pushing forward without sufficient consultation with stakeholders. Teachers, parents, and human rights organizations have raised various concerns about AIDT. In June 2024, a National Assembly public petition demanding the "Suspension of the Ministry of Education's 2025 AI Digital Textbook Implementation" received 53,884 signatures and was referred to the National Assembly's Education Committee. The main points of the petition called for verification of whether AIDT are an effective educational method, citing concerns about:

- The pros and cons of digital device-based learning identified during remote learning due to COVID-19
- Students' over-dependence on smart devices and related harmful effects
- Delays in textbook development and certification schedules, and opposition from teachers

Subsequently, on August 28, 2024, 127 organizations, including the Korean Teachers and Education Workers Union, National Parents' Network for Better Education, and Politically Engaged Mothers, formed the 'Joint Committee Against AI Digital Textbooks.' They criticized the Ministry of Education for rushing to become the world's first country to implement AIDT at a national level, arguing that this means Korean children could be the first to experience unprecedented negative effects. From September 6 to October 1, 2024, the Joint Committee conducted a signature campaign demanding the suspension of AIDT implementation and the formation of a public deliberation committee to review policy validity. According to reports, more than 100,000 citizens participated in the signature campaign within a month.

3.4.2. Concerns About the Educational Effectiveness of AIDT

If the educational benefits of AIDT were clear, there would be no reason for teachers and parents to oppose it. However, opposing groups worry that AIDT could negatively impact youth by increasing digital device usage time, potentially leading to smart device addiction. They also point to research showing that digital learning methods can reduce literacy compared to traditional methods. While "personalized education" is emphasized, AIDT are actually structured to present tasks repetitively, raising concerns that they might become "personalized rote learning." Therefore, more thorough examination is needed to determine if they are appropriate educational tools for supporting student growth and development.

However, the Ministry of Education is pushing forward with implementing the system while ignoring the education sector's demands for verification of AIDT's' effectiveness. Despite requests to introduce the system gradually by implementing it partially as a pilot project and evaluating the results, the Ministry has announced plans to introduce it without sufficient preparatory phases. Starting March 2025, mathematics, English, and information technology subjects will be implemented for all students in grades 3-4 of elementary school, first year of middle school, and first year of high school, while Korean

language will be introduced only for students requiring special education.

3.4.3. Legal Basis for AIDT Implementation

The Ministry of Education claims to have established a legal basis for introducing AIDT by amending the enforcement decree "Regulations on Textbooks." However, the National Assembly Research Service published a report stating that introducing AIDT through an enforcement decree amendment violates the constitutional principle of education system legalism.

Considering the education sector's resistance, the opposition party proposed an amendment to the Elementary and Secondary Education Act that would classify AIDT as 'educational materials' rather than 'textbooks,' allowing schools to use them as needed and implement them gradually. This amendment passed the National Assembly plenary session on December 26, 2024. However, the Ministry of Education requested the government to exercise its 'veto power,' arguing that the policy implementation is only meaningful when AIDT have the legal status of 'textbooks.' The controversy continues, with the National Assembly holding an "AIDT Verification Hearing" on January 17, 2025.

3.4.4. Issues from a Personal Information Protection Perspective

3.4.4.1. Legal basis for Personal Information Collection and Use in AIDT

AIDT necessarily collect various types of student personal information as they provide learning guidance based on each student's individual characteristics and achievement levels. Such collection and use of personal information must be conducted on a legal basis. Typically, schools' collection of student personal information such as personal details, academic status, and attendance records is based on the Elementary and Secondary Education Act (Article 25).

The publishers of AIDT are the personal information controllers responsible for handling personal information in the operation of AIDT. However, since publishers lack legal grounds for processing personal information under education-related laws, they must obtain consent from the data subjects. This consent must be freely given (Article 17 of the Enforcement Decree of the Personal Information Protection Act), meaning that data subjects should be able to refuse consent if they wish. If consent is effectively forced at schools, it cannot be considered legally valid consent. If AIDT are used as 'textbooks' as planned by the Ministry of Education, they would necessarily apply to all students in certain schools, which inevitably conflicts with the principle of freely given consent. Therefore, for AIDT to be smoothly implemented, there needs to be a priority on establishing legal grounds through social consensus regarding the processing of personal information in the use of AIDT.

3.4.4.2. Need for Clarification of Personal Information Flow and Controllers in AIDT Usage

Various institutions and systems are involved in the use of AIDT, including schools (teachers), the Education Administration Information System (NEIS), AIDT services, integrated gateways, and learning data hubs. In this process, it is unclear which personal information controllers (responsible entities) are handling what personal information and under what legal basis. There is a need to clarify the flow of personal information between different institutions, specific personal information items, and legal grounds. This clarity is necessary to prevent personal information breaches and establish clear accountability when problems occur.

3.4.4.3. Excessive Personal Information Collection and Surveillance of Students

AI digital textbooks collect not only basic student information but also very detailed information including:
- Login/logout times, frequency, and intervals in the system
- Number of learning content items studied
- learning time and patterns
- Number of posts
- Assessment results
- Question answering activity
- Assignment status
- Learning history

The monitoring extends beyond learning processes in school to after-school learning activities, which amounts to monitoring students' entire lives and risks seriously infringing on students' privacy and autonomy. Such surveillance can create pressure on students and is undesirable from an educational perspective as well.

3.4.5. Concerns Regarding Artificial Intelligence Functions

The European Union's AI Act prohibits the use of 'AI systems that infer emotions in workplaces and educational institutions.' It also classifies systems that evaluate educational levels or monitor prohibited student behavior during exams as 'high-risk' and imposes strict obligations. These obligations include:
- Establishing risk management systems
- Training data evaluation and management systems
- Creation and maintenance of technical documentation
- Providing detailed information to users (deployers)
- Post-implementation monitoring measures
- Maintaining log records
- Security measures
- Accessibility measures

In Korea, along with the amended Elementary and Secondary Education Act that invalidated AIDT, the AI Framework Act also passed the National Assembly on December 26, 2024. According to this Act, 'AI that evaluates students in early childhood, elementary, and secondary education' is also classified as high-impact AI. However, compared to the EU AI Act, the obligations of high-impact AI providers are not specifically defined in the law.

While there are no legal standards for the requirements of AI-related functions in AIDT, as previously examined, the "AI Digital Textbook

Development Guidelines" present considerations for establishing data management policies, including personal information, and principles for AI risk management. The technical review process uses the 'Self-Technical Verification Report' submitted by developers as baseline material. Several issues can be raised regarding this:

First, it is questionable whether the guidelines sufficiently include all considerations related to AIDT development and operation. For example, they do not address operational requirements such as maintaining technical documentation and log records related to development and operation processes, monitoring, or the Ministry of Education's supervision mechanisms.

Second, there is a lack of transparency regarding what AI functions are included in the AIDT that will be actually used, what AI technologies are applied, and how the guidelines' principles are implemented. For example:
- Whether generative AI technology is used in the AI tutor function
- How bias and hallucination issues are controlled
- If emotion recognition AI is used to detect students' emotions
- Whether the accuracy of emotion recognition AI can be trusted

While the content of textbooks is publicly available, allowing social evaluation of the certification process, the technical aspects of AIDT are difficult to assess through the published AIDT. Without this information being public, it is difficult to build social trust in whether the guidelines' standards are being properly followed. When a freedom of information request was submitted to the Ministry of Education regarding these matters, they only provided a general response stating that "this is a matter where each developer autonomously develops textbooks according to the requirements of the guidelines."

3.4.6. Deterioration of Local Education Finances

The Ministry of Education is mandating the use of AIDT alongside existing printed textbooks. Consequently, each school must incur additional costs to purchase AIDT. Unlike printed textbooks, which can be used for more than a year after a single purchase, AIDT require monthly subscription payments to publishers. According to the National Assembly Research Service's estimate of AIDT subscription fees for elementary, middle, and high schools, an annual budget of 406.7 billion won will be needed in 2025. By 2028, when AIDT are introduced for all subjects and grades, they project an astronomical annual budget requirement of 1.734 trillion won. However, the Ministry of Education has not established separate funding measures for introducing AIDT, leaving metropolitan and provincial offices of education to bear all these burdens through local education finances. Investing such massive funds in AIDT will inevitably lead to budget cuts in essential projects such as improving aging school facilities, multicultural education, and enhancing basic academic skills. The Joint Committee Against AIDT criticizes that pouring enormous taxpayer money into unproven AIDT is not only excessive waste of public education funds but also only serves to benefit the edutech industry.

3.4.7. International Community's Concerns About the World's First AIDT

Education International (EI) held its 10th World Congress in Buenos Aires, Argentina, from July 29 to August 2, 2024. EI is the world's largest teachers' organization, representing 32 million teachers from 383 education unions across 178 countries. At the 10th Congress, the EI Executive Board adopted a resolution on "Technology, Artificial Intelligence and the Future of the Teaching Profession." The resolution expresses strong concerns about the indiscriminate implementation of AI and digital technology in educational settings. Randi Weingarten, EI Executive Board member and President of the American Federation of Teachers (AFT), who proposed the resolution, stated that "AI technology has been indiscriminately entering the education sector since COVID-19." She pointed out several issues:

- Digital divide among students
- Deepening inequality
- Public education funds flowing to private companies

She called for stronger international response at the EI level. This contradicts Education Minister Lee Joo-ho's statement during the National Assembly Education Committee's 'Education Issues Inquiry,' where he claimed that "international organizations, and even teacher organizations like EI, expressed opinions that (AI digital textbooks) are extremely effective."

Jeon Hee-young, Chairperson of the Korean Teachers and Education Workers Union (KTU) who attended the congress as both a KTU representative and EI executive board member, expressed strong support for the resolution. She also pointed out several issues arising from Korea's AIDT implementation:

- Conflicts with the legal concept of textbooks
- Digital rights violations and personal information leakage concerns
- Insufficient verification of potential impacts on students

Additionally, she urgently requested EI to dispatch an international investigation team composed of relevant experts to verify and respond to Korea's AIDT Initiative. EI indicated that they would positively consider this request.

Chapter 4

Current Status of Artificial Intelligence in Social Welfare

4.1. Introduction

Various countries are incorporating digital technologies into their welfare systems, and recently these services have evolved to utilize artificial intelligence (AI). Governments are promoting data-driven welfare by using AI and big data to identify welfare blind spots and provide personalized services without human intervention. However, data-driven welfare services raise several concerns, including privacy invasion, excessive control, unfair decisions, and potential accountability issues. Furthermore, while the welfare sector primarily emphasizes human care as its most crucial aspect, delegating these responsibilities to AI poses significant risks. This transition could potentially marginalize vulnerable populations from protection or amplify their psychological insecurities.

4.2. AI-Related Regulations in Social Welfare

Despite the widespread use of AI in the public welfare sector, there are no specific laws governing its implementation. The closest regulatory frameworks are the artificial intelligence ordinances or AI industry promotion and support ordinances that have been arbitrarily established by local governments. However, these local government ordinances have limitations in that they have narrow scope and application, and they tend to focus more on AI development and industry promotion rather than serving as comprehensive regulations for AI implementation.

4.3. AI Systems for Vulnerable Population Care and Support

There is no official documentation regarding what types of AI are being implemented, where they are being used, or to what extent in the social welfare sector. It appears that local governments are actively adopting these systems, though this can only be inferred through media releases and promotional materials. We will examine cases of AI systems that are most visibly being utilized for vulnerable population care and social welfare purposes.

4.3.1. Current Status

AI systems are being introduced into healthcare programs for vulnerable populations, particularly the elderly. In 2023, the Ministry of Health and Welfare, Korea Social Security Information Service, and Korea Health Promotion Institute jointly announced the 'AI and IoT-based Senior Health Management Project.' The project aims to transform the existing

system of health center workers making home visits into a technology-based remote health monitoring service, shifting from in-person visits to virtual care. The target population consists of 'seniors aged 65 and above who require management of frailty and chronic conditions, as well as improvements in health behaviors.' The project provides these individuals with various devices including wrist activity trackers, Bluetooth-enabled scales, blood pressure monitors, blood glucose meters, and both standard and display-equipped AI speakers. These devices are used to monitor and screen various health indicators.

Additionally, Naver, a private corporation, has launched and begun expanding its 'Clova Care Call' service, where AI systems make phone calls to check on individuals' health status and whether they have eaten their meals.

When participating in the AI and IoT-based Senior Health Management Project, participants are provided with various devices for health monitoring, including wrist-band biosensors, Bluetooth-enabled scales, blood pressure monitors, and blood glucose meters, which are used to measure their physical data. Typically, the data collected through these devices is recorded via mobile apps that are synced with the devices. If abnormal readings are detected or if no measurements are taken for more than a week, staff are required to follow up by phone. The user's information is shared with assigned healthcare professionals, including nurses, nutritionists, and exercise specialists.

For those who don't have smartphones or struggle with using measurement devices, AI speakers (available in both display and non-display versions) are provided as an alternative. These systems monitor users' wellbeing through phone calls or direct conversations via the speakers. Similarly, Naver's Clova Care Call service monitors users' overall health through interactive conversations. The AI maintains context across conversations by remembering previous interactions about medication adherence, meal patterns, and other health-related topics.

4.3.2. Problems

4.3.2.1. Sensitive Data Leaks and Privacy Violations

The AI and IoT-based Senior Health Management Project necessarily stores not only personal identification information but also personal health data collected through devices, such as blood pressure, heart rate, and blood glucose levels. Additionally, with consent, information related to health status including medication usage, disease diagnoses, and hospitalization frequency is stored and transmitted. This information is highly sensitive as it can reveal an individual's medical history and requires even stronger protection than regular personal information. While such data should be collected and protected only to the minimum extent necessary for the project's purpose, there is a need to verify whether the data collection is indeed being minimized appropriately for the stated purpose of caring for vulnerable populations.

In the Guidelines for Adoption and Use of Large Language Models announced by the government in April 2024, 'deriving insights by analyzing data such as age, region, medical conditions, and occupation

of medical benefit recipients for recipient management' is presented as a future AI application area. This essentially proposes the use of sensitive information for AI development. Such biometric-based profiling could qualify as high-risk AI. According to the European Union's AI Act, AI systems that could pose significant risks to human rights should be classified as high-risk and subject to strict operator obligations.

However, South Korea currently lacks such mandatory regulations. Additionally, there are concerns about algorithmic opacity and tendencies to collect excessive information, while the purposes and processing procedures of how collected health information is linked to AI systems and transmitted to medical staff remain unclear. In particular, a significant issue is that these systems are being implemented without social welfare-specific AI regulations in place, leading to a lack of established principles and procedures for the safe utilization of high-risk AI.

4.3.2.2. Concerns Regarding Personal Data Consent Practices

There may be issues regarding clear understanding and consent for device usage among vulnerable populations, particularly the elderly, who are the target service recipients. As digitally vulnerable groups often have low technological literacy, they might consent without properly understanding the significance and implications of the data they're providing. Therefore, safeguards are needed, such as implementing additional procedures to thoroughly explain terms and conditions and verify recipients' comprehension levels.

Clova Care Call service requires users to consent to both the collection and sharing of their health (sensitive) information with third parties; without this consent, the service cannot be used. For the Clova Care Call service, local governments including district offices serve as the primary personal information controllers, while Naver, the service operator, is designated as a third party (separate personal information controller) authorized to receive personal information. However, if local governments are the main operators of the Clova Care Call service, even with Naver's operational involvement, this should be considered merely a 'commissioned' processing, where Naver should only process personal information for purposes delegated by local governments. The reason for defining Naver's access as third-party provision rather than commissioned processing appears to be for Naver's independent use of sensitive information for 'improving call quality and AI service quality through AI learning', as stated in the <Health (Sensitive) Information Collection, Use, and Provision Consent Form>.

The Personal Information Protection Act requires separate consent for data collection and third-party provision, and services cannot be denied based on refusal to consent to uses beyond the primary purpose (Article 22). However, Clova Care Call's consent form combines consent for necessary personal information and 'AI learning for service improvement' into a single agreement, making service contingent upon this combined consent, which may violate the Personal Information Protection Act. Unlike commissioning data to specialized institutions, Clova Care Call's third-party provision fundamentally differs as information could

be used for purposes beyond service provision. Such forced consent practices should be avoided.

In projects involving private companies alongside government agencies, most cases involve sharing collected data with companies through third-party provision consent. This raises concerns about collected data being used for commercial purposes or shared with additional third parties like insurance companies or pharmaceutical firms, potentially leading to issues such as increased insurance premiums. Furthermore, if data security is compromised through hacking or inadequate protection, sensitive information could be leaked, resulting in privacy violations.

4.3.2.3. Consideration of Social Impact of AI Systems for Vulnerable Population Care and support

In cases where individuals either do not consent to the AI and IoT-based Senior Health Management Project or cannot use it due to the digital divide, they continue to receive traditional in-person care visits. However, these individuals are likely to be assigned to remaining staff after personnel have been allocated to AI-based services, rather than having dedicated caregivers. This could result in increased vulnerability for those who prefer traditional care methods. Furthermore, from the perspective of public officials working in care services, AI implementation may not reduce their existing workload but rather replace it with different tasks, and they might face potential staff reductions.

The introduction of AI systems inevitably reduces opportunities for human-to-human interaction. Increased reliance on technology for health management can decrease interactions between family members and caregivers. While AI speakers (both with and without displays) provided alongside measurement devices primarily deliver care by gathering health information through conversations and experiencing various interactions, AI-based care systems have both advantages and drawbacks. They may weaken emotional stability among the elderly and damage human bonds that cannot be replaced by technology. Research shows that elderly individuals may attempt to interact with care robots, perceive them as human-like, and even feel companionship with them. However, the same research indicates that in cases of severe depression and loneliness, relationships with robots can turn into obsessive attachments, potentially leading to self-isolation and other negative impacts on social relationships. Therefore, the implementation of such systems requires a cautious approach and should be preceded by thorough consideration of their effects on both workers and users.

4.4. Social Welfare System

4.4.1. Current Status

4.4.1.1. Identification and Management of Welfare Recipients

The Ministry of Health and Welfare had been providing welfare services through a system where local government social workers would conduct initial phone consultations with households suspected to be in crisis, followed by in-depth consultations and home visits to connect them with

social security benefits or private services. However, in July 2024, as part of the '4th Welfare Blind Spot Discovery' project, AI systems were introduced at the initial consultation stage. This project was expanded nationwide in November of the same year.

The purpose of introducing AI was to delegate initial consultations to AI systems, allowing government officials to focus on intensive consultations with households in crisis, thereby expediting the identification and support of welfare emergencies. Participating local governments first send automated text messages informing residents about upcoming consultation calls. Subsequently, the AI system conducts initial consultations with previously identified at-risk households to assess their need for welfare assistance. The content of these AI-conducted initial consultations is automatically provided to local government officials through the system and used as reference material for in-depth consultations and household visits.

In addition to local governments, other public institutions have also implemented AI check-in call services using the previously mentioned 'Clova Care Call'. The service makes AI-powered calls to elderly pension recipients and parents of educational staff who have pre-registered, checking on their health, meals, sleep, and exercise while engaging in casual conversation. According to Clova Care Call, the system can provide personalized conversation services by remembering previous interactions, enabling delicate emotional care. The service also includes care monitoring features such as connecting users to emergency services (119) or hospitals when signs of health-related crises are detected during calls.

4.4.1.2. Detection of Fraudulent Welfare Claims

While national support funds are essential for protecting and helping the vulnerable, they consistently face controversies regarding fraudulent claims. There is an urgent push to implement AI systems that use data tracking to address this issue. In South Korea, the Ministry of Economy and Finance has introduced a system that uses algorithms to detect suspicious patterns, such as tax invoice cancellations after issuance and transactions between family members. Similarly, Korea Electric Power Corporation has incorporated AI technology to identify cases where electricity bill discounts for vulnerable groups are being improperly claimed. This AI application has replaced the manual process of monthly verification of households' eligibility for discounts.

Additionally, the Social Security Information Service has implemented an electronic voucher fraud detection system for social services. The system is designed to detect major irregular payment patterns for each target program, such as batch payments, duplicate payments, late-night transactions, and consecutive payments. When fraudulent use is suspected, payments are suspended pending verification of legitimacy.

4.4.2. Problems

4.4.2.1. Data Bias and Lack of Verification

According to an Amnesty International report, Serbia introduced an automated data-driven system for determining support eligibility when implementing its social card registration system in 2022, but this

resulted in harm to marginalized populations. The registration system built individual socioeconomic profiles by pulling data such as income, age, household composition, health status, and employment status from existing government databases. While not fully automated, as social workers were required to review and classify the database entries, the system particularly disadvantaged marginalized groups whose government database records were often outdated. Amnesty International pointed out that "(this system) has reduced the economic realities of people living off informal workstreams and with hugely varying personal circumstances to, often outdated, data points."

In 2010, France implemented an algorithmic system that assigns risk scores to detect potential welfare fraud. While the system regularly updates these scores using individual and family data, it was discovered that the criteria for higher risk scores included parameters that discriminate against vulnerable populations. These discriminatory factors included low income, unemployment status, living in disadvantaged neighborhoods, allocating a large portion of income to rent, and working with disabilities.

These case studies highlight the critical importance of input data quality in AI systems. AI systems learn and make decisions based on their training data, which often reflects existing societal biases. When the input data contains inherent biases related to geographic location, income levels, or ethnicity, the AI is likely to perpetuate these biases in its decision-making, potentially worsening inequality and further disadvantaging vulnerable populations.

The Serbian example shows that even with human oversight, when AI implementation results in reduced staffing and human role is limited to validating AI decisions, identifying unfairness becomes challenging. Moreover, deploying insufficiently validated algorithms in social welfare can have devastating consequences for marginalized groups who depend on immediate financial support for survival. This necessitates robust monitoring and supervision systems.

Furthermore, it's crucial to ensure transparent communication with affected individuals about the implementation of AI in welfare services.

4.4.2.2. Recipients' Rights to AI Decision Explanation and Appeals

Welfare-related algorithms are classified as high-risk AI systems that require human rights impact assessments and regulatory oversight. As AI systems make decisions based on data-driven algorithms, they often struggle to comprehend unique individual circumstances and contexts. A case from Serbia illustrates this limitation: a woman lost her welfare recipient status when donations of 20,000 Serbian dinars (around 170 euro) received for her suddenly deceased daughter's funeral expenses were counted as regular income. This example demonstrates AI's inability to properly interpret complex situations such as temporary financial changes or family circumstances.

In such cases, citizens have the right to demand clear explanations about why their benefits were revoked, understand how these decisions were made, and appeal incorrect judgments. However, when AI makes these decisions, obtaining proper explanations can become extremely challenging.

Vulnerable populations, in particular, may struggle to protect their rights as they often find it difficult to understand complex administrative procedures and technical terminology. In Serbia's case, people were unable to receive explanations from social workers about why their benefits were terminated, and they had to navigate complicated procedures, visiting various government offices to file appeals and resolve their situations. The right to explanation and appeal is fundamental for citizens to protect their rights, and this is especially crucial for vulnerable populations.

Additionally, partially automated systems require careful attention. As seen in the U.S. COMPAS case, even though judges reviewed the results, they could ultimately rely on AI judgments. Given the potential for automation bias and dependency, it's essential to ensure that affected individuals have sufficient opportunities to appeal decisions.

In Korea, when using AI systems to identify welfare recipients or detect fraudulent claims, similar problems to those experienced abroad could arise. Therefore, policies and systems need to be established to identify and control these risks.

Chapter 5

Artificial Intelligence Framework Act of Korea

5.1. Introduction

On December 26, 2024, the 'Framework Act on Artificial Intelligence Development and Building Trust (hereinafter "AI Framework Act")' passed the National Assembly. Even amid political turmoil, including President Yoon Suk-yeol's sudden declaration and lifting of martial law and his subsequent impeachment by the National Assembly, the AI Framework Act was packaged as 'legislation for people's livelihood' and passed without partisan disagreement. While civil society demanded the inclusion of regulations to control AI risks in the Framework Act, the government, ruling and opposition parties, and major media outlets agreed that more supportive policies were needed to enhance Korea's AI industry competitiveness. The interim report of the "International scientific report on the safety of advanced AI" pointed out the risk that global competition for AI development could lead to deregulation in various countries, with Korea being a prime example of such concerns. With the passage of the AI Framework Act, discussions surrounding the bill appear to be concluding for now. However, as the rapid development and implementation of AI technology is likely to cause various problems, discussions to amend the AI Framework Act to address these risks may begin soon. This paper aims to summarize the main contents, progress, and issues of the AI Framework Act.

5.2. Main Contents of the AI Framework Act

The main contents of the AI Framework Act that passed the National Assembly are as follows:

a. The purpose of this Act is to protect citizens' rights and dignity and contribute to improving their quality of life and strengthening national competitiveness by stipulating matters necessary for the sound development of artificial intelligence and establishing a foundation of trust (Article 1).

b. This Act defines artificial intelligence, high-impact artificial intelligence, generative artificial intelligence, AI ethics, and AI business operators (Article 2).

c. The Minister of Science and ICT shall establish and implement a Basic Plan for Artificial Intelligence every three years to promote AI technology and industry and strengthen national competitiveness, following deliberation and resolution by the National Artificial Intelligence Committee. The Basic Plan must include matters concerning the basic direction of AI policy, nurturing of professional personnel, and establishing a foundation of trust (Article 6).

d. A National Artificial Intelligence Committee shall be established under the President to deliberate and

decide on major policies regarding the promotion of the AI industry and building a foundation of trust for AI. The Committee shall deliberate and decide on matters concerning the establishment of basic plans, promotion of AI utilization, and regulation of high-impact AI (Articles 7 and 8).

e. The Minister of Science and ICT may designate an AI Policy Center for the development of AI-related policies and the establishment and dissemination of international norms, and may operate an AI Safety Research Institute to ensure AI safety (Articles 11 and 12).

f. The government may support projects such as domestic and international trend surveys, institutional research, technology commercialization, and research and development to facilitate AI technology development and ensure safe and convenient use. The Minister of Science and ICT may promote projects such as establishing standards for AI technology standardization (Articles 13 and 14).

g. The Minister of Science and ICT may nurture relevant professional personnel and promote various measures to secure international expertise for the development of AI technology and promotion of the AI industry (Article 21).

h. The national and local governments may promote functional, physical, and regional clustering of companies, institutions, or organizations conducting research and development of AI and AI technologies to enhance the competitiveness of the AI industry and AI development and utilization (Article 23).

i. The government may establish and announce AI ethical principles that include matters such as safety and reliability, accessibility, and contribution to human life and prosperity to promote AI ethics. The Minister of Science and ICT shall establish implementation measures for AI ethical principles and must publicize, promote, and educate about these measures (Article 27).

j. The Minister of Science and ICT may promote projects to support voluntary verification and certification activities conducted by corporations, institutions, organizations, etc., to ensure AI safety and reliability (Article 30).

k. AI business operators providing products or services using high-impact AI or generative AI must :

- Notify users in advance of such use

- Indicate when outputs are generated by generative AI when providing generative AI or products/services using it

- When providing virtual outputs that are difficult to distinguish from reality using AI systems, Clearly notify or indicate this fact to users (Article 31).

l. AI business operators must implement measures such as risk identification, assessment, and mitigation to ensure the safety of AI systems whose cumulative computation used for training exceeds the standards prescribed by Presidential Decree (Article 32).

m. When providing high-impact AI or products/services using such AI, AI business operators must implement measures to ensure safety and reliability (Article 34).

n. If the Minister of Science and ICT discovers or becomes aware of suspected violations of this Act, they may require AI business operators to submit materials or have government officials conduct necessary investigations. If violations are confirmed, the Minister may order necessary measures to stop or correct the violations (Article 40).

5.3. Civil Society's Response to the AI Framework Act

Since the 21st National Assembly (May 30, 2020 - May 29, 2024), several assembly persons including Lee Sang-min, Jung Pil-mo, and Yoon Young-chan have proposed artificial intelligence bills. On February 14, 2023, the Science and ICT Committee's Legislative Review Subcommittee of the 21st National Assembly suddenly passed an artificial intelligence bill. Despite being a newly enacted law, there had been insufficient public hearings or social discourse regarding the risks of AI. The subcommittee meeting minutes reveal that the bill was hastily passed without substantial debate on key issues. Moreover, while this legislation claimed to be a framework law for AI, it focused solely on industrial promotion, including the principle of "permitt first, regulate later," and lacked provisions for controlling AI risks or imposing responsibilities on business operators. Civil society criticized the bill passed by the Legislative Review Subcommittee and demanded that it not be processed by the Science and ICT Committee.

The Ministry of Science and ICT (MSIT) was effectively leading the content of the AI bill in the National Assembly. On April 26, 2023, the MSIT sent a response to the opinion paper submitted by civil society to the National Assembly's Science and ICT Committee, and civil society subsequently submitted a rebuttal opinion.

The National Human Rights Commission of Korea's (NHRCK) opinion on the AI bill was also completely ignored during the National Assembly's deliberation process. On August 21, 2023, the NHRCK recommended to the Speaker of the National Assembly that the 'permit first, regulate later' principle should be removed from the AI bill, and provisions should be established to prevent and regulate human rights violations and discrimination, including human rights impact assessments.

Globally, awareness of the need to control AI risks was growing, and countries began establishing administrative and legislative frameworks to regulate AI. On October 30, 2023, the Biden administration issued an Executive Order 14110, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." On December 8, 2023, the European Union reached an agreement on the "AI Act," becoming the first in the world to comprehensively regulate high-risk AI.

While civil society agreed that South Korea needed legislation to regulate AI risks, they urged that the AI bill pending in the National Assembly should be discarded due to its excessive focus on industrial promotion, and called for substantial discussions in the 22nd National Assembly. The MSIT claimed it had modified the AI bill to reflect civil society's opinions, but it had merely removed the 'permit first, regulate later' clause without incorporating demands to strengthen business operators' responsibilities. In fact, during closed-door meetings with companies, the Ministry revealed that it had "minimally incorporated civil society's opposing views." Although the Ministry pressured for the passage of the AI bill using the Seoul AI Summit as an excuse, the bill ultimately failed to pass during the 21st National Assembly and was discarded due to the expiration of its term.

As soon as the 22nd National Assembly opened, AI bills were continuously proposed. Civil society submitted an opinion paper to the National Assembly outlining the direction for establishing AI legislation and held public discussions. Furthermore, given AI's society-wide impact, they urged the formation of a 'National Assembly Special Committee on AI' to conduct reviews across standing committees.

Assembly person Jung Jeom-sik's bill was effectively the government and ruling party's proposal, signed by all members of the People Power Party. As promised by the MSIT at the end of the 21st National Assembly, it did not include the 'permit first, regulate later' clause. However, the rest of its content was not significantly different from the consolidated bill (passed by the Science and ICT Committee's Legislative Review Subcommittee) of the 21st National Assembly. As pointed out in the civil society opinion paper, the bill had several limitations: ▲it did not specify which AI systems should be prohibited, ▲it narrowly defined high-risk AI areas, ▲it had insufficient provisions regarding high-risk AI operators' (developers and deployers) obligations and lacked penalty clauses, limiting its effectiveness. Additionally, ▲it lacked provisions for the rights and remedies of people affected by AI, and ▲it excluded obligations for general-purpose AI operators, such as disclosure of training data. Civil society questioned whether the MSICT, with its bias toward industrial promotion, was suitable as the primary ministry responsible for AI.

On September 24, 2024, the Science and ICT Committee of the 22nd National Assembly held a committee-level public hearing on AI bills. Civil society submitted detailed opinion papers addressing major issues in the AI bills proposed to the National Assembly. However, the Committee passed the consolidated AI bill (AI Framework Act) that merged multiple AI bills after only two review subcommittee meetings. As soon as the 22nd National Assembly opened, similar AI bills focusing on deregulation were proposed one after another, and 19 of these bills were consolidated and processed.

Despite some improvements compared to the government and ruling party's proposal, the Committee's passed version was not significantly different from the consolidated bill of the 21st National Assembly. It still ▲lacks provisions specifying which AI systems should be prohibited, and ▲has insufficient penalty provisions for high-risk AI operators' violations of their duties. Fines are only imposed when operators fail to comply with the MSIT's corrective orders. ▲While it fortunately includes a definition of persons affected by AI, it fails to specify their rights and remedies. ▲It also does not include provisions for general-purpose AI operators' obligations, such as disclosure of training data. Moreover, it added a problematic clause: ▲it excludes AI systems for defense or national security purposes from the law's application.

Civil society separately prepared and proposed a 'Civil Society's AI Framework Act' that could guarantee public safety, human rights, and democracy by preventing AI-induced risks and establishing remedial procedures for when problems occur. However, this proposal failed to be introduced as a bill. As entirely new legislation and a framework law, the AI Framework Act should have reviewed civil society's proposals regarding AI risk regulation and rights remedies, and undergone sufficient public

discourse. To use it as a stepping stone for continuous improvement of the AI Framework Act in the future, civil society submitted their prepared bill as a legislative petition on December 3, 2024, introduced by Democratic Party Assembly person Kim Nam-geun.

5.4. Issues with the AI Framework Act

Civil society criticizes that the AI Framework Act passed by the National Assembly focuses excessively on industrial promotion while ignoring human rights. Nevertheless, there are some improvements compared to the initially proposed government and ruling party's bill, which include the following points.

- The 'permit first, regulate later' principle that was included in some proposed bills was not incorporated into the final AI Framework Act.
- The Act adopts the OECD's definition of 'AI systems' to ensure international consistency and includes the concept of 'affected persons' in its definitions.
- High-impact [high-risk] AI now explicitly includes 'student assessment in early childhood, elementary, and secondary education.'
- The basic principles (Article 3) establish the 'rights of affected persons' to receive clear and meaningful explanations about the reasons and logics behind AI decisions, within technically and reasonably feasible bounds, particularly "when AI significantly impacts people's lives, physical safety, and fundamental rights."
- The Act grants stakeholders, including affected persons, the right to file reports and complaints.
- It gives the MSIT new authority to investigate legal violations through fact-finding missions and issue corrective orders.
- Transparency requirements have been strengthened, including mandatory disclosure and labeling obligations for generative AI such as deepfakes.
- Although a weak provision, the Act introduces an obligation for high-impact AI operators to 'make efforts' to conduct impact assessments.

However, the Act largely excluded key regulations that civil society has been demanding. The main issues are as follows:

First, there are no provisions regarding AI systems that should be prohibited. The government opposed such provisions, arguing that no countries except the European Union have established regulations banning certain AI systems, and that such bans could hinder industrial development.

Second, the scope of high-impact AI remains narrow compared to the EU AI Act. For instance, it's unclear whether the following are included:
- AI systems used for analyzing and utilizing biometric information in areas other than criminal investigation and arrest
- AI systems used in the exercise of state authority that could infringe on fundamental rights, such as investigation and prosecution
- AI systems used for emotion recognition
- AI systems used by the judiciary or executive branch for judgments, decisions, and adjudications
- AI systems used in the operation of information and communications networks
- AI systems used to influence elections, voting behavior, and voting results
- AI systems that could affect product safety

Third, the penalties for violations of high-impact AI operators' obligations are inadequate. Not only are the operators' obligations themselves lacking in specificity, but there are no direct penalties for violations. Fines are only imposed when operators fail to comply with corrective orders issued by the MSIT regarding such violations.

Fourth, while it is fortunate that the definition of persons affected by AI has been included, the Act notably lacks provisions regarding their rights and remedies.

Fifth, the Act does not include obligations for general-purpose AI operators, such as the disclosure of training data. This could potentially conflict with personal information protection and copyright protection.

Sixth, the effectiveness of the AI human rights impact assessment is questionable as it only imposes an 'obligation to make efforts.' While the National Human Rights Commission's release of an 'AI Human Rights Impact Assessment Tool' in May 2024 is a positive development, the MSIT, as the primary governing body, has been ignoring this initiative.

Seventh, the Act includes a new toxic provision that exempts AI systems used for defense or national security purposes from its application. While the government and National Assembly claim these should be regulated by separate legislation, it remains uncertain when such laws will be enacted. Instead of this exemption, the AI Framework Act could have covered defense and national security AI systems while allowing for the creation of special laws specifically for these purposes if necessary. This provision was included at the request of the National Intelligence Service (NIS). Given the NIS's long history of human rights violations and political interference, along with the military's recent involvement in martial law, there are serious concerns about how to control potential misuse of AI technology by these institutions.

* End *