THE FUTURE OF FREE SPEECH

THAT VIOLATES MY POLICIES

AI LAWS, CHATBOTS, AND THE FUTURE OF EXPRESSION

Directed by
Jordi Calvet-Bademunt,
Jacob Mchangama, and Isabelle Anzabi

OCTOBER 2025

Acknowledgments

The Future of Free Speech is an independent, nonpartisan think tank based at Vanderbilt University. Our mission is to reaffirm freedom of expression as the foundation of free and thriving societies through actionable research, practical tools, and principled advocacy. We envision a world in which the right to freedom of expression is safeguarded by law and strengthened by a culture that embraces diverse viewpoints.

This project was led by Jordi Calvet-Bademunt (Senior Research Fellow), Jacob Mchangama (Executive Director), and Isabelle Anzabi (Research Associate) at The Future of Free Speech. Together, they also drafted the chapters on the European Union and the United States of America.

We are grateful to Justin Hayes, Director of Communications, for overseeing the design of the report; Wendy H. Burch, Chief Operating Officer, for coordinating all administrative aspects of the project; and Sam Cosby, Director of Development, for leading the funding efforts that made this work possible.

We extend our thanks to the leading experts who contributed chapters on their respective jurisdictions: Carlos Affonso Souza (Brazil), Ge Chen (China), Sangeeta Mahapatra (India), and Kyung Sin (K.S.) Park (Republic of Korea). We are also grateful to Kevin T. Greene and Jacob N. Shapiro of Princeton University for their chapter, "Measuring Free Expression in Generative AI Tools."

We thank all the experts who contributed to individual chapters of this report; their names are listed in the relevant sections.

We are further indebted to Barbie Halaby of Monocle Editing for her careful editorial work across all chapters, and to Design Pickle for the report's design.

Finally, we are especially grateful to the Rising Tide Foundation and the Swedish Postcode Lottery Foundation for their generous support of this work, and we thank Vanderbilt University for their collaboration with and support of The Future of Free Speech.

Publication Date: October 2025

Preface

In this report, we explore the ways in which public and private governance of generative artificial intelligence (AI) shape the space for free expression and access to information in the 21st century.

Since the launch of ChatGPT by OpenAI in November 2022, generative AI has captured the public imagination. In less than three years, hundreds of millions of people have adopted OpenAI's chatbot and similar tools for learning, entertainment, and work.[1] Anthropic, another AI giant, now serves more than 300,000 business customers.[2] AI companies are valued in the hundreds of billions of US dollars[3], while established technology giants such as Google, Meta, and Microsoft are investing billions in the race to dominate the field.[4]

Generative AI refers to systems that create content — including text, images, video, audio, and software code — in response to user prompts.[5] Chatbots such as ChatGPT are the most visible examples, but generative AI is rapidly being embedded into the tools people use every day for both communication and access to information, from social media and email to word processors and search engines.

Recognizing generative AI's potential for expression and access to information, The Future of Free Speech undertook a first-of-its-kind analysis of freedom of expression in major models. In February 2024, we assessed the "free-speech culture" of six leading systems, focusing on their usage policies and responses to prompts.[6] Our findings revealed that excessively broad and vague rules often resulted in undue restrictions on speech and access to information.[7] By April 2025, when we updated this work, we observed signs of change: Some models showed greater openness.[8]

This current report builds on those foundations and pursues a more ambitious goal. Supported by leading experts, The Future of Free Speech undertakes a deeper examination of how national legislation and corporate practices shape freedom of expression in the era of generative AI. "*That Violates My Policies": AI Laws, Chatbots, and the Future of Expression* explores:

- AI legislation in Brazil, China, the European Union, India, the Republic of Korea, and the United States.[9] In this report, AI legislation refers to laws and public policies addressing AI-generated content, with

1 MacKenzie Sigalos, "OpenAI's ChatGPT to Hit 700 Million Weekly Users, Up 4x from Last Year," *CNBC*, August 4, 2025, https://www.cnbc.com/2025/08/04/openai-chatgpt-700-million-users.html.
2 Hayden Field, "Anthropic Is Now Valued at $183 Billion," *The Verge*, September 2, 2025, https://www.theverge.com/anthropic/769179/anthropic-is-now-valued-at-183-billion.
3 Kylie Robison, "OpenAI Is Poised to Become the Most Valuable Startup Ever: Should It Be?," *Wired*, August 19, 2025, https://www.wired.com/story/openai-valuation-500-billion-skepticism/; Krystal Hu and Shivani Tanna, "OpenAI Eyes $500 Billion Valuation in Potential Employee Share Sale, Source Says," *Reuters*, August 6, 2025, https://www.reuters.com/business/openai-eyes-500-billion-valuation-potential-employee-share-sale-source-says-2025-08-06/.
4 Blake Montgomery, "Big Tech Has Spent $155bn on AI This Year: It's About to Spend Hundreds of Billions More," *The Guardian*, August 2, 2025, https://www.theguardian.com/technology/2025/aug/02/big-tech-ai-spending.
5 Cole Stryker and Mark Scapicchio, "What Is Generative AI?," IBM Think, March 22, 2024, https://www.ibm.com/think/topics/generative-ai.
6 Jordi Calvet-Bademunt and Jacob Mchangama, *Freedom of Expression in Generative AI: A Snapshot of Content Policies* (Future of Free Speech, February 2024), https://futurefreespeech.org/wp-content/uploads/2023/12/FFS_AI-Policies_Formatting.pdf.
7 Calvet-Bademunt and Mchangama, *Freedom of Expression in Generative AI*.
8 Jordi Calvet-Bademunt, Jacob Mchangama, and Isabelle Anzabi, "One Year Later: AI Chatbots Show Progress on Free Speech — But Some Concerns Remain," *The Bedrock Principle*, April 1, 2025, https://www.bedrockprinciple.com/p/one-year-later-ai-chatbots-show-progress.
9 To select the countries, we considered Stanford University's 2023 Global AI Vibrancy Ranking (the most recent available at the time of writing), along with factors such as geographic diversity, population size, democratic and freedom status, and the presence of existing or emerging AI-related legislation.

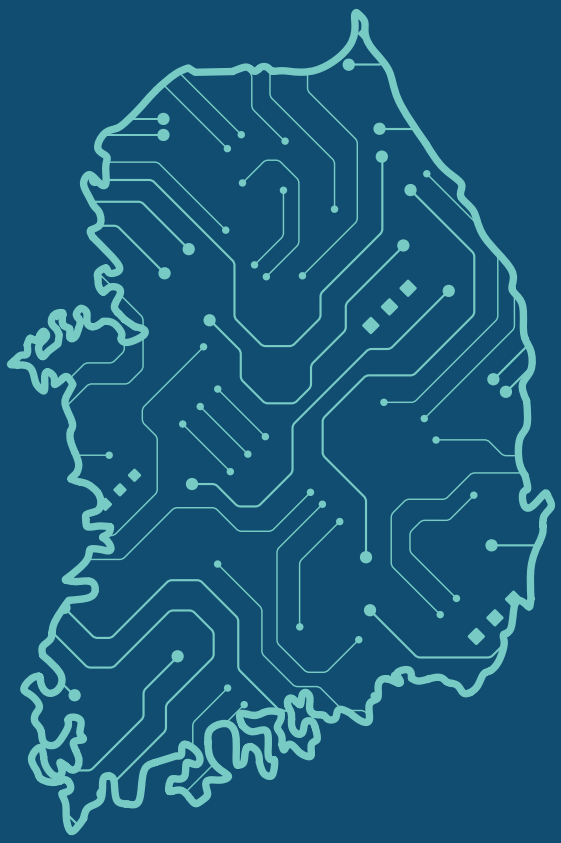
particular focus on elections and political speech, hate speech, defamation, explicit content (including child sexual abuse material and nonconsensual intimate images), and copyright. We also consider measures that actively promote freedom of expression, such as AI literacy initiatives and policies supporting cultural and linguistic diversity.

- Corporate practices of major AI developers, including Alibaba, Anthropic, Google, Meta, Mistral AI, DeepSeek, OpenAI, and xAI.[10] We examine their usage policies, model performance in responding to prompts, and the limited available information on their training data and development processes.

This report seeks to provide a rigorous and timely analysis of how generative AI is reshaping the space for free expression in both the public and private spheres. Building on these insights, The Future of Free Speech is developing guidelines to help policymakers and companies ensure that generative AI protects and enhances freedom of expression and access to information, two cornerstones of democratic societies.

In an era of rapid technological change, safeguarding free expression is a matter not only of rights but of preserving the conditions for open, informed, and thriving democracies.

10 We selected major models from leading companies that are accessible through a web interface and include text-generation capabilities. In addition, we considered the geographic location of the model provider and the degree of openness of the models.



Artificial Intelligence and Freedom of Expression in the Republic of Korea

Kyung Sin (K.S.) Park*

* Professor, Korea University Law School; AB in physics, Harvard University; JD, UCLA Law School; visiting professor at the Law Schools of UCLA, UC Irvine, and UC Davis; director, Open Net; co-founder, Open Net Korea; former commissioner of Korea Communications Standards Commission. Park has written academically and been active in internet, free speech, privacy, defamation, copyright, and artificial intelligence. Internationally, he served on the Global Network Initiative board and the High Level Panel of Legal Experts on Media Freedom and currently serves as an advisor to Freedom Online Coalition. Park also was a key drafting partner of international standards on online free speech and privacy: namely, Principles of Application of International Law on Communication Surveillance and International Principles on Intermediary Liability.

Abstract

South Korea has fallen behind other developed countries in protecting freedom of speech. As artificial intelligence can be used to make speech, and speech or other access to knowledge is needed to make artificial intelligence, the generally depressed state of freedom of speech in this country has thereby suppressed the freedom to make or use artificial intelligence for the purpose of speech or access to knowledge. To illustrate the double-edged effects of free speech on AI, the stringent application of defamation laws has suppressed online speech, including defamatory material made with artificial intelligence. Additionally, the general unavailability of court decisions due to the threat of liability under truth defamation laws and data protection laws is hampering people's AI-mediated access to legal knowledge.

Also, AI itself has been the target of regulation through the AI Basic Act, fashioned after the EU's AI Act. With this act, market-facing activities using AI are subject to transparency and safety mitigation obligations even before creation and sharing, as well as to administrative agencies' control after the fact, which will be denser for high-impact AI, generative AI, and deepfakes. Because AI replaces the decision-making and reasoning part of a human action, imposing such substantive and procedural obligations on that part is constitutionally allowed only when proportional to the magnitude of risk that such decision-making and reasoning poses. The AI Basic Act tries to enact such proportionality between regulation and danger in the text of the law, but does not seem to succeed all the time. On a separate note, the Sex Crimes Special Punishment Act and the Elections Act practically ban use of deepfakes in sexual material and election material featuring another person without their consent. It is doubtful that such laws achieve the proportionality they aim for, since they do not even require falsity as an element.

Copyright law and data protection law need to be analyzed separately as they can potentially regulate the act of training either as "copying" or "data-processing" and overzealous application of those laws can restrict the scope of training data. Whether machine learning on and pseudonymization of the training data constitutes "fair use" or personal data processing has not been reviewed by the courts or decided by administrative agencies in any conclusive manner. Such uncertainty will generate chilling effects on AI training efforts.

Kyung Sin (K.S.) Park
Professor, Korea University Law School; AB in physics, Harvard University; JD, UCLA Law School; visiting professor at the Law Schools of UCLA, UC Irvine, and UC Davis; director, Open Net; co-founder, Open Net Korea; former commissioner of Korea Communications Standards Commission. Park has written academically and been active in internet, free speech, privacy, defamation, copyright, and artificial intelligence. Internationally, he served on the Global Network Initiative board and the High Level Panel of Legal Experts on Media Freedom and currently serves as an advisor to Freedom Online Coalition. Park also was a key drafting partner of international standards on online free speech and privacy: namely, Principles of Application of International Law on Communication Surveillance and International Principles on Intermediary Liability.

1. Introduction

This chapter explores South Korea's legal landscape governing AI, focusing on legislation, policies, and case law that intersect with freedom of expression. It also examines related areas such as copyright and defamation laws that impact AI's role in society.

South Korea's National Strategy for Artificial Intelligence, released in December 2019, lays out a roadmap for advancing into the top tier of global AI leaders by 2030.[1] The strategy emphasizes four pillars: building robust AI infrastructure (including data platforms and high-performance computing), boosting R&D (AI semiconductors, foundational technologies), ensuring regulatory flexibility, and nurturing a new generation of AI start-ups.[2] To grow human capital, the strategy aims to expand AI literacy across all age groups, integrate AI education in the military, public, and private sectors, and build lifelong learning infrastructures.[3] In parallel, the government launched "human-centered AI ethics standards" in 2020 — a voluntary code encouraging inclusivity, transparency, explainability, and accountability in AI development.[4] These principles are designed to foster public trust and socially responsible AI adoption.

A major legislative milestone was the passage of the AI Basic Act by the National Assembly on December 26, 2024; it was promulgated on January 21, 2025, and will go into effect on January 22, 2026. This national-level law, only the second of its kind globally (after the EU AI Act), unifies 19 prior bills and establishes both promotional measures and regulations. However, the overall approach is deemed a "permit-first-and-regulate-later" type, which is the same approach that the South Korean government has typically taken toward new technologies.[5]

South Korea continues to advance its AI agenda, first through an informal director-level public-private engagement in April 2024,[6] and then through the National AI Committee, an inter-ministerial entity established in September 2024 and chaired directly under the president, which now serves as a central policy control tower integrating government and industry input.[7]

The new president, Lee Jae-Myung, has stayed on course with the AI initiatives of the previous regime, all the way down to the motto "Become the World's Big Three" (United States, China, and South Korea). He has promised to invest a higher percentage of government spending than other developed countries in AI and to induce more than 100 trillion KRW (a little less than USD 100 billion) of private investment, build national AI data centers equipped with more than 50,000 GPUs (competing with the US Stargate project), create regional AI industrial clusters, increase the availability of government data for AI training purposes, provide financial

1 Ministry of Science and ICT, *National Strategy for Artificial Intelligence*, December 2019, https://www.msit.go.kr/bbs/view.do?sCode=eng&nttSeqNo=9&bbsSeqNo=46&mId=10&mPid=9.
2 Digital Watch Observatory, "The National Strategy for Artificial Intelligence of South Korea," October 2019, https://dig.watch/resource/the-national-strategy-for-artificial-intelligence-of-south-korea.
3 Digital Watch Observatory, "National Strategy."
4 Korea Information Society Development Institute, *National Guidelines for AI Ethics*, December 2020, https://ai.kisdi.re.kr/eng/main/contents.do?menuNo=500011.
5 Digital Watch Observatory, "Overview of AI Policy in 10 Jurisdictions," December 2024, http://v45.diplomacy.edu/updates/overview-of-ai-policy-in-10-jurisdictions.
6 Ministry of Science and ICT, "Korea Establishes the High-Level Consultative Council on Artificial Intelligence Strategy as the Top-Level Governance Structure for AI," press release, April 2024, https://www.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=4&mPid=2&pageIndex=&bbsSeqNo=42&nttSeqNo=994&searchOpt=ALL&searchTxt=.
7 National AI Committee, https://aikorea.go.kr/web/main.do.

support for the development of Korean native neural processing unit (NPU) chips, and provide financial support for large language models (LLMs) that everyone in Korea can use for free.[8]

In sum, the Korean government's AI strategy is development-oriented across the political spectrum, while the legislature has responded with an EU-style law using risk-based due process; however, its regulatory strength is in doubt.

8 Business Korea, "Korean Government to Invest $11.56 Billion in AI Infrastructure Over Next 5 Years," June 19, 2025.

2. Substantive Analyses

2.1. General Standards of Freedom of Expression

Given that AI can be used to make speech, the state of freedom of speech is expected to affect the freedom to use artificial intelligence. For instance, heightened sanctions on defamation will apply also to defamatory photos or new articles made with AI. But even more primarily, current and future versions of artificial intelligence are made by machine learning on the data available to AI developers, so the existing and upcoming freedom of speech protections will affect the freedom to create AIs. For instance, the general unavailability of court decisions due to the threat of liability under truth defamation laws and data protection laws is hampering people's AI-mediated access to legal knowledge through artificial intelligence. It is therefore important to assess a general state of freedom of speech to gauge the level of freedom bestowed upon artificial intelligence.

In South Korea, freedom of speech is enshrined in the Constitution but operates within a legal environment shaped by the nation's unique historical, geopolitical, and cultural context.

The Constitution of the Republic of Korea guarantees freedom of speech under Article 21:

(1) All citizens shall enjoy freedom of speech and the press, and freedom of assembly and association.

(2) Licensing or censorship of speech and the press, and licensing of assembly and association shall not be recognized...

(4) Neither speech nor the press shall violate the honor or rights of others or undermine public morals or social ethics.[9]

Article 21 is supposed to align with international human rights standards, notably Article 19 of the International Covenant on Civil and Political Rights (ICCPR), which South Korea ratified in 1990. However, Article 21(4) has been interpreted in a way that embraces serious limitations concerning the protection of honor, rights, and public morals, providing the legal basis for restrictions that are broader than those found in some other liberal democracies.[10]

Most notably, the National Security Act (NSA), enacted in 1948, criminalizes acts deemed to benefit anti-state organizations, notably North Korea. Article 7 penalizes the praise and encouragement of such organizations' activities.[11] The NSA has been repeatedly challenged but upheld by the Constitutional Court, which argues

9 Constitution of the Republic of Korea, art. 21, https://elaw.klri.re.kr/eng_service/main.do.
10 Freedom House, *Freedom on the Net: South Korea* (2023), https://freedomhouse.org/country/south-korea/freedom-net/2023.
11 Amnesty International, "Freedom of Expression in South Korea: A Continuing Challenge" (2019), https://www.amnesty.org/en/documents/asa25/0022/2019/en/.

its necessity due to the unique security situation on the Korean Peninsula. In its 1990 decision 89Hun-Ga113, this court emphasized the need for strict interpretation to minimize infringement on constitutional rights, and Article 7 was amended to include as an extra element of the crime "knowledge of a threat to the nation's existence, security, and liberal democratic order," implying that it was adapting the "clear and present danger" test. However, human rights bodies have continued to criticize the NSA as a tool to suppress dissent and freedom of expression.[12]

South Korea's Criminal Act contains provisions that penalize both true and false statements if made solely to defame another (Article 307) and insults not based on fact (Article 311).[13] The Constitutional Court upheld these provisions in 2017 Hun-Ma 1113 (2021) and in 2020 Hun-Ba 456 (2020), using the personality right as a value to be protected from statements that are true or that constitute mere opinions. In contrast, the UN Human Rights Committee specifically advised that truth shall be an absolute defense to the claims of defamation.[14] Also, South Korea handles a large volume of criminal prosecutions,[15] some of which were filed to protect the reputation of high-level officials such as President Suk-Yeon Yoon.[16]

South Korea has enacted a mandatory notice-and-takedown regime, in which online platforms have the legal obligation to take down illegal content upon notice from the rightsholders. Under this regime, even many lawful postings have been taken down.[17] South Korea also established online administrative censorship whereby the administrative agency Korean Communication Standards Commission deliberates on specific online content and issues the blocking or takedown requests to the local internet service providers (ISPs) or platforms when it is "necessary for nurturing sound communication ethics."[18] Thus, many web pages constituting legitimate civic discourse under international human rights standards were taken down or blocked.[19]

South Korea has had a European-style data protection law since 2011, since upgraded to obtain the European Commission's adequacy decision under the General Data Protection Regulation (GDPR).[20] Data protection laws such as GDPR and the Korean law entitle all data subjects to limited control about data about them (namely "personal data"), and all processing of personal data is restricted by various requirements, both before and after publishing or sharing, which are also exempt under publicly recognized situations (i.e., contract enforcement, public interest, life and safety, the data controller's overwhelming interest). Despite the derogation under GDPR in favor of freedom of expression, the Korean data protection law did not institute a strong derogation but has only added an extraneous criminal provision[21] that seems to take away the balance carefully built into the main consent-related provisions between the need for use of data and the protection of data subjects.[22]

12 Constitutional Court Decision 2010Hun-ba70, June 28, 2012.
13 Kyung S. Park and Jong-Sung You, "Criminal Prosecutions for Defamation and Insult in South Korea with a *Leflarian* Study in Election Contexts," *University of Pennsylvania Asian Law Review* 12 (2017), https://scholarship.law.upenn.edu/alr/vol12/iss3/4.
14 UN International Covenant on Civil and Political Rights, "Concluding Observations on the Fourth Periodic Report of the Republic of Korea," CCPR/C/KOR/CO/4, November 3, 2015.
15 Park and You, "Criminal Prosecutions."
16 US State Department, *2023 Country Reports on Human Rights Practices: South Korea*, https://www.state.gov/reports/2023-country-reports-on-human-rights-practices/south-korea/.
17 Kyung Sin Park, "From Liability Trap to the World's Safest Harbour: Lessons from China, India, Japan, South Korea, Indonesia, and Malaysia," in *Oxford Handbook of Online Intermediary Liability*, ed. Giancarlo Frosio (Oxford University Press, 2020), 251–76.
18 Kyung Sin Park, "Administrative Internet Censorship in Korea," *Soongsil Law Review* 3 (January 2015): 91–115.
19 Open Net, "International Coalition to Support Filing of a Suit to Stop South Korea's Shutdown of Womenonweb.kr," March 13, 2022, https://www.opennetkorea.org/en/wp/3547.
20 European Commission, *Decision on the Adequate Protection of Personal Data by the Republic of Korea with Annexes*, December 17, 2021, https://commission.europa.eu/document/e9453177-f192-4416-a147-3c57adc468c4_en.
21 Kyoungmi Oh, "Regrettable Court Ruling That Filing a Police Complaint Violates Personal Information Protection Act," Open Net, November 7, 2024, https://www.opennetkorea.org/en/wp/6072; for an academic treatment, see 박경신 [Kyung Sin Park], 공익적 언사와 개인정보보호법 [Public interest speech and data protection law] 법학연구 (경상국립대학교 법학연구소) [Legal Studies (National Kyung Sang University)] Vol, 33, no. 1 (2025) pp. 25–50 (Korean only).
22 Kyung Sin Park, "Data as Public Goods or Private Properties? A Way Out of Conflict Between Data Protection and Free Speech," *UC Irvine Journal of International, Transnational, and Comparative Law* 6 (2021): 77.

In sum, South Korea's freedom of speech is moderately suppressed by the substantive criminal defamation and insult laws, a deficient safe harbor regime for online intermediaries, administrative censorship, and restrictive data protection law, which will apply equally when the speech is made with AI and will reduce substantially the training data available for the development of AI. AI technologies, such as chatbots and content generators, can produce content violating any of these laws. Under current laws, individuals or entities deploying such AI systems could be held liable for these regulations, even if the content produced or shared was not intentionally harmful. For instance, generative AI can produce texts that are inadvertently false and negatively affect another's reputation. This potential liability may lead to self-censorship and hinder the development and use of AI technologies that facilitate expression.

Now, there is no AI-specific law, regulation, or precedent applying the general rules of defamation, national security, or data protection to AI-generated content. In the main body of this chapter, we consider AI-specific laws that apply the defamation-type norms more severely to AI-generated contents (i.e., deepfakes) used in electoral contexts or sexual contexts. But first, we will look at the AI Basic Act, which imposes obligations on the *application* of AI to various uses. Also, we'll discuss the current controversies on copyright law and data protection law that directly restrict the machine learning processes, which potentially constitute communicative activities protected under Article 21 of the Korean Constitution.

2.2. AI-Specific Legislation and Policies

In December 2024, South Korea's National Assembly passed the AI Basic Act, which consolidated all of the previous legislative initiatives aimed at regulating AI. Potentially marking a significant step in AI governance, the AI Basic Act adopts a risk-based approach, categorizing AI systems based on their potential impact on human life and rights, as the EU AI Act does, down to the prominent national defense and security exception.[23] "High-impact" AI systems, particularly those used in critical sectors like health care and public decision-making, are subject to stricter regulations in terms of explainability, safety, accountability, and transparency. High-impact AI is defined as "AI system that can possibly cause material impact or danger to human life, physical safety and basic rights" operating in a number of areas, such as energy, potable water, health care, digital health care, nuclear energy, biometric identification for criminal investigation or arrest purposes, hiring, loan applications, transportation, public benefit eligibility, and primary and second education.

However, some argue that the similarities may be superficial: Unlike the EU AI Act, there are no provisions regarding prohibited artificial intelligence practices; the penalties for violations of obligations are inadequate (i.e., a fine up to KRW 30,000,000, roughly equivalent to USD 27,000); and there is no provision for a remedy for "those affected by AI."[24]

From a free speech perspective, AI is at the far end of a spectrum of automation that human civilization has been traversing from its beginning. Or one may say it is at a possible pinnacle of that trend, considering that, after automating agriculture (e.g., harvest machines), transportation (e.g., automobiles), computation (e.g., computers), and a long list of human activities, we are finally attempting to automate thinking or decision-making itself. Automating otherwise innocuous human activities is always met with regulation, as automation always amplifies the inherent risk in the activity being automated. Driving an automobile is regulated by

23 For comparison with the EU AI Act, see Hosuk Lee-Makiyama, Jimmyn Parc, and Claudia Lozano, "Korea's New AI Law: Not a Progeny of Brussels," ECIPE, https://ecipe.org/blog/koreas-new-ai-law-not-brussels-progeny.
24 Oh Byung-Il, "South Korea's AI Framework Act Enactment Biased Toward Industry Growth," Association for Progressive Communication, March 2025, https://www.apc.org/en/blog/south-koreas-ai-framework-act-enactment-biased-toward-industry-growth.

a licensing scheme, which was justified by the fact that — unlike walking, running, or bicycling — motored mobility amplifies the risk of injury to oneself and others. What is being automated by AI? Decision-making or thinking. What is the inherent risk associated with decision-making or thinking? It depends on what human activity the automated decision-making or thinking is applied to. If the automated decision-making is applied to driving an automobile, then the risk inherent in motored mobility may be intensified; for instance, automobiles will crowd streets without human controllers. However, is that a risk inherent in driving or a risk inherent in thinking?

Thinking, imagining, feeling, loving, and other "mental actions" belong to the domain of human activity that has been protected under freedom of expression and freedom of opinion for the very reason that these mental actions do not cause harm, according to philosopher John Stuart Mill's harm principle. Should automation of thinking be subject to a new restriction just as automation of moving was subjected to regulations? On what grounds? We have not yet imposed any regulation on the use of software in writing, drawing, painting, communicating, signaling, or other communicative actions, even though the use of software does amplify and magnify whatever communicative or informational harms such actions may present. On what grounds do we suddenly impose such regulation because the power of automation is delivered through LLMs as opposed to non-LLM software? It is from this foundation that we can evaluate the AI Basic Act.

2.2.1. Applied Only to Market Activities

The AI Basic Act applies only to "AI businesses": the corporations, associations, individuals, or state agencies that "conduct business," either by "developing and providing AI" or by "using AI to provide AI goods or AI services" (Article 2, item 7). On one hand, this means that AI development itself is not affected by the law. On the other hand, because "AI goods" and "AI services" are defined by whether AI was used in development, manufacturing, production, or distribution (Article 2, item 6), a very broad spectrum of all goods and services will be affected by the act. As a relevant example, if a journalist working for a commercial media outlet uses ChatGPT to embellish a news article, it will be an "AI good" and thus subject to the law. Contrarily, a casual YouTube creator clearly not conducting "business" who uploads videos made with generative AI will not be subject to the law.

In summary, only the entities providing something available in the market for goods and services will be subject to the AI Basic Act. Meanwhile, as the use of AI spreads to various decision-making processes within businesses, a very broad spectrum of goods and services provided by those entities will be subject to the act.

For the purpose of free speech, this is significant because freedom of expression also protects research or other "internal" activities taking place as part of the back-office operation, and the AI Basic Act is not regulating these activities. The AI Basic Act, unlike the EU AI Act, does not provide an explicit exemption for scientific research or premarket testing. Neither does it provide for regulatory sandboxes (EU AI Act, Article 53). However, because the initial scope is limited to market-facing activities, these exemptions and sandboxes are not necessary, as the implication is that such internal activities are not governed by the AI Basic Act.

There are other differences with the EU AI Act that do not directly concern freedom of expression.[25]

25 Lee-Makiyama et al., "Korea's New AI Law": "Nor does it explicitly single out general-purpose AI with distinct obligations … While Korean [institutional] users 'should prioritise' [using] systems that have been tested and certified (article 30) for high-impact AI use-cases, the Act does not explicitly require the use of such systems, unlike the EU AI Act … Both Korean and EU laws require ex-ante assessments of the higher categories of high-impact or high-risk. In the AI Basic Act, high-impact AI systems must undergo an *ex-ante* review submitted to the Ministry of Science and ICT, and an expert committee can be established to advise if necessary … However, unlike in the EU, the AI Basic Act does not require third-party conformity assessment of high-risk systems and the technical

2.2.2. Transparency and Safety Obligations and "High-Impact AI"

Exactly what transparency and safety obligations will be imposed depends on whether generative AI, high-impact AI, or deepfake have been used, to which heightened transparency and safety obligations will be applied. We can evaluate the impact of these more stringent obligations on freedom of speech.

For transparency, AI businesses "providing goods or services that use high-impact or generative AI" must notify in advance the users that they are based on said AI (Article 31, para. 1). AI businesses "providing generative AI or the goods or services based thereon" must label on the results of the AI the fact that they were generated by the generative AI (para. 2). AI businesses "using any type of AI to provide virtual sound, image, video or other products difficult to distinguish with the reality" — i.e., deepfakes — must notify or label them so the user clearly knows that the results were created with AI, provided that the notification or labeling on the artistic or creative expressions can be done in a manner that does not encumber display or enjoyment of the results of AI (para. 3). Note that this notification or labeling obligation applies only when generative AI, high-impact AI, or deepfake is involved.

For freedom of expression, such notification or labeling obligation constitutes "compelled speech," as the creator of certain goods or services must disclose the fact of having used AI. For instance, the Microsoft Office 365 suite provides Copilot (generative AI) as a service to suite users whenever they operate any of the included applications. This means that any journalist, academic author, or creative writer using Copilot to produce their output or provide their service will have to label these or notify customers of that. Failure to notify will be penalized with a fine of KRW 30,000,000, while no penalty is stipulated for failure to label (Article 43). To the extent that freedom of expression includes the right to speak anonymously, in any language, or even in code incomprehensible to some (e.g., encryption),[26] it is not clear how the notification/labeling obligation would be justified.

For safety, AI businesses using above a certain cumulative compute for training must secure the safety of the AI system by "identifying, evaluating and mitigating risks throughout the life of the AI" and by "establishing a risk management system that monitors and responds to any safety incidents related to AI" (Article 32, para. 1). They also must report the results of their safety-related efforts to the science ministry (Article 32, para. 2). Speech deserves restriction when the speech is likely to cause an external harm, i.e., "a clear and present danger." However, the requirement that all AI businesses operating above certain compute limits must take and report on safety measures even though the government has not produced any reason to believe AI presents a clear danger seems incongruent with the concept of due process. Although there is no penalty for failing to comply, a failure in this regard can be the basis for civil liability.

documentation requirements are less rigid under Korean law. Overall, the EU AI Act outlines a structured pre-market conformity assessment (article 43-51) which *de facto* is a licensing regime, whereas the Korean law emphasises *post-hoc* oversight supported by new agencies (e.g. AI Safety Research Institute, article 12) that is similar to antitrust enforcement … The fines [KRW 30 million, equivalent to less than USD 2,700] are just a fraction of the fines in the EU that can amount to €35 million or 7% of the total worldwide annual turnover. Furthermore, systems that are compliant with the Basic AI Law cannot be held accountable for civil liabilities, whereas EU opens for civil liabilities under the AI Liability Directive with reverse burden of proof — i.e. where developers are assumed to be liable without any proof of the opposite … Korea avoids the most restrictive and binding ex-ante interventionist approaches for these 'high impact' activities and does not impose [burden-shifting like] strict product liability for AI developers."

26 David Kaye, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, UN A/HRC/29/32, May 2015, https://documents.un.org/doc/undoc/gen/g15/095/85/pdf/g1509585.pdf.

There are other obligations imposed on high-impact AI. AI businesses providing goods or services based on high-impact AI are responsible for the following (Article 34, para. 1):

1. Establishing and administering a risk-management plan

2. Establishing and implementing a plan to explain within the technologically possible extent the final result, the standard used in arriving at the result, the description of the training data used, etc.

3. Establishing and administering a user protection plan

4. Human management of high-impact AI

5. Drafting and storing the documents evidencing the safety and reliability measures taken

6. Other measures to guarantee the safety and reliability of high-impact AI as resolved by the National AI Committee.

In addition, the science ministry can issue advisories about any of the above (Article 34, para. 2).

In another provision, all AI businesses providing AI or goods or services based thereon must evaluate in advance and, if necessary, check with the science ministry about whether their goods and services constitute "high-impact" AI (Article 33). There is no penalty for lack of or error in such evaluation, but it is problematic that the obligation is imposed on all AI businesses providing AI or goods or service based thereon. The justification seems to be that use of AI by itself is somehow deemed so dangerous that such businesses must have a government entity determine in advance whether they should be categorized as high-impact AI.

Again, what AI is doing is computation, an activity ordinarily left in the domain of freedom of expression. Just to repeat the argument above, before LLMs, we might reference 20 books to write an article. After LLMs, we are referencing a billion books to write that same article — though, of course, most of the books referenced will be found not to be useful and will be discarded for the purpose of writing that article. But AI will replace the labor and experience needed to identify 20 useful books so that even a complete novice or a 10-year-old child could write an article of similar caliber. Other than that a broader swath of the public can engage in the activities previously limited to an elite group, AI does not import any inherent danger.

The duty to disclose to the government automation of an AI business' market-facing activities (or risk penalties in the event they turn out to be high-impact AI) effectively forces submission to the government of all AI businesses. It is not clear how such encumbrance on automated thinking will be justified. It does not amount to prior restraint since government approval is not an explicit requirement for use of AI in goods and services, but it is unusual among other countries' policies. One reason that prior restraint (in American jurisprudence) or prior censorship (in European jurisprudence) is heavily frowned upon is because such a system forces one to disclose to the government one's otherwise confidential thoughts and opinions.[27]

27 Carly Nyst, "Two Sides of the Same Coin — The Right to Privacy and Freedom of Expression," Privacy International, February 2018, https://privacyinternational.org/blog/1111/two-sides-same-coin-right-privacy-and-freedom-expression.

To the extent that "high-impact AI" is defined to target only those AI usages possibly involving material impact on human rights and safety, the impact on freedom of speech seems generally contained. It is like a proverbial law that penalizes "speech that creates a clear and present danger of substantive evils." For instance, if AI is applied to an inherently dangerous area such as nuclear energy, the public need to regulate nuclear energy still remains and therefore the composite act of applying artificial intelligence to nuclear energy needs be regulated. The areas of human activities most enumerated for high-impact AI are already covered by heavy area-specific regulations, and if the decision-making part of that specific activity in those areas is automated by artificial intelligence, it seems reasonable for one more layer of regulation for the very reason that the decision-making is executed by machines, not humans.

2.2.3. Domestic Representative

All AI businesses domestic or foreign (Article 4) are subject to the Korean AI Basic Act if they "affect the domestic market or users." Some commentators believe this is a broader extraterritorial reach than that of the EU AI Act,[28] which covers "(a) providers placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the Union, irrespective of whether those providers are established or located within the Union or in a third country... [and] (c) providers and deployers of AI systems that have their place of establishment or are located in a third country, where the output produced by the AI system is used in the Union (Article 2)."

Important for our purposes, all AI businesses above a certain numbers of users or amount of revenue without address or place of business within Korea are required to appoint a domestic representative (Article 36) at the penalty of a fine up to KRW 30,000,000 (Article 43). The international human rights community has embraced a similar domestic representative requirement on data controllers because of the risk of privacy violation implicated in personal data processing. However, appointing a domestic representative for the mere use of AI in a businesses does not seem to be justified by any such risk — unless automated decision-making by itself is considered risky. Note that there is no such appointment requirement for domestic AI businesses.

From a free speech perspective, requiring a foreign AI business to have a domestic representative abolishes the company's right to anonymous communication affecting the South Korean market or its users.

2.2.4. Administrative Control

Under Article 40 of the AI Basic Act, the Korean science ministry is empowered to investigate businesses that it suspects of breaching any of the following requirements:

- Labeling for generative AI outputs (Article 31, para. 2) or labeling/notification for deepfakes (Article 31, para. 3);

- Implementation of safety measures and submission of compliance results for AI systems exceeding computational thresholds set by Presidential Decree (Article 32, paras. 1 and 2); and

- Adherence to safety and reliability standards for high-impact AI systems (Article 34, para. 1)

28 Park Kwang-bae and Sakshi Shivahare, "South Korea's New AI Framework Act: A Balancing Act Between Innovation and Regulation," Future of Privacy Forum, April 2025, https://fpf.org/blog/south-koreas-new-ai-framework-act-a-balancing-act-between-innovation-and-regulation/.

When potential breaches are identified, the science ministry has the authority to carry out necessary investigations, including to conduct on-site investigations and to compel AI businesses to submit relevant data. If violations are found, the ministry can issue corrective orders, requiring businesses to immediately halt noncompliant practices and implement necessary remediation measures.

To the extent that use of AI is a communicative activity, the highest judicial courts of many states have consistently held that administrative bodies restricting communicative activities without the safeguard of judicial review amounts to prior restraint, which violates the principle of freedom of expression.[29] Here, the science ministry may issue such orders to stop the communicative activity.

Now, free speech is not absolute; it can be regulated by administrative bodies under certain conditions. Safety measures by definition are directed at "substantive evils," not the speech itself, and the labeling requirements do not directly block speech.

2.3. Defamation

Regarding the general state of freedom of expression in Korea, criminal defamation law, "truth defamation" law, and insult law are vigorously prosecuted. However, no precedent so far signals that AI-generated contents are more severely prosecuted under these laws. South Korea does have "deepfake" laws concerning explicit content or electoral contexts, which are intended to restore the reputation or honor of the person whose facial data are nonconsensually used in the deepfakes.

2.4. Explicit Content

The proliferation of deepfake technology has emerged as a significant concern in South Korea, particularly regarding nonconsensual, sexually explicit content. In response, the government has enacted laws criminalizing the creation, distribution, and even possession of deepfake pornography, with penalties including imprisonment and substantial fines.[30]

While these measures aim to protect individuals from harm, they also raise questions about their impact on freedom of expression. The text of Article 14-2 of the Sexual Crimes Special Punishment Act (Distribution of False Video Products) follows:

> (1) A person who edits, synthesizes, or fabricates (hereafter referred to as "edits, etc.") in this Article) photograph, video, or audio (hereafter referred to as "photograph, etc." in this Article) featuring the face, body or voice of a person in a form that may cause sexual desire or shame against the will of the person who is the subject of the video, etc., shall be punished by imprisonment with labor for not more than 7 years or a fine of not more than 50 million won. [Amended October 16, 2024]
>
> (2) A person who distributes, etc. an edited, synthesized or fabricated material (hereafter referred to as "edited material, etc." in this Article) (including a duplicate of its duplicate; hereinafter the same applies in

29 E.g., Bantam Books, Inc. v. Sullivan, 372 U.S. 58 (1963); Little Sisters Book & Art Emporium v. Canada (Minister of Justice), 2000 SCC 69 (2000) (Can.); Rappler, Inc., Petitioner v. Andres D. Bautista, Respondent, [2016] PHSC 85 (Hong Kong); Poland v. Parliament and Council, 62019CJ0401 (EU); Disini v. The Secretary of Justice, [2014] G.R. No. 203335 (Philippines); French Constitutional Court — Decision n 2009-580 DC of 10 June 2009 (only in French, June 10, 2009); French Constitutional Court — Decision n 2020-801 DC of 18 June 2020; Turkish Constitutional Court, nos. 2014/149 (October 2, 2014, annulling the law), followed by no. 2014/3986 (April 2, 2014, lifting Twitter.com ban), no. 2014/4705 (May 29, 2014, lifting YouTube.com ban).
30 Hyung-Jin Kim, "In South Korea, Deepfake Porn Wrecks Women's Lives and Deepens Gender Conflict," AP News, October 2024, https://apnews.com/article/south-korea-deepfake-porn-women-df98e1a6793a245ac14afe8ec2366101.

this Article) under paragraph (1), or a person who distributes the edited material, etc. against the will of the person thus featured, etc., afterwards even if it is not contrary to the will of the person featured in the video material, etc. at the time of editing, etc. under paragraph (1), shall be punished by imprisonment with labor for not more than 7 years or by a fine not exceeding 50 million won. [Amended October 16, 2024]

(3) A person who commits a crime under paragraph (2) by means of information and communications networks against the will of the person subject to video works, etc. for the purpose of making profits shall be punished by imprisonment with labor for a limited term of not less than 3 years. [Amended October 16, 2024]

(4) A person who possesses, purchases, stores, or views an edited material, etc., or its duplicates referred to in paragraph (1) or (2) shall be punished by imprisonment with labor for not more than 3 years or by a fine not exceeding 30 million won. [Newly inserted October 16, 2024]

(5) A person who habitually commits any of the crimes provided for in paragraph (1) through (3) shall be aggravatingly punished by up to 1/2 of the punishment for each crime. [Newly inserted May 19, 2020; October 16, 2024]

Previously, any deepfake material composed of the face of one person with the exposed body of another person would be treated as defamation against the first person since it is considered a visual statement that attributes to the first person a defamatory situation that has not taken place. With the new provision in the Sexual Crimes Special Punishment Act, production and distribution of deepfakes of another person without his or her consent were considered a different crime, and the penalty has been made stronger, increasing from five years under criminal defamation to seven years under the Sexual Crimes Special Punishment Act.

The broad scope of these laws may inadvertently suppress legitimate uses of deepfake technology, such as satire or artistic expression. The crime of defamation typically requires that a reasonable person may believe the defamatory statement to be true. Therefore, satire or other patently false statements would not be considered defamation because a reasonable person will not believe it to be true. Under the Sexual Crimes Special Punishment Act, however, there is no such defense. The law would apply even to a composite of someone's face with a sexually desirous or shameful situation in a clearly nonrealistic way such that no one would believe that person to have engaged in that situation. For instance, sexual material involving government officials' nudity will be punished.

Even more important, there is no requirement that the deepfake mislead viewers about whether the person featured engaged in that sexual situation. For instance, the truthful representation of a sexual situation will still be prosecuted if the color of the sky, completely irrelevant to whether the event took place, was edited, synthesized, or fabricated. For that matter, the law does not require the photo to mislead viewers in any way. If the photo was edited in a way to remove a visual hindrance or change the lighting to show the underlying event more clearly — and therefore truthfully — the photo can be still prosecuted, even if the editing does not attempt to generate or enhance the sexually desirous or shameful nature of the material.

The guidelines of the Sexual Crimes Special Punishment Act work in parallel to the preexisting provision about the recording or photographing of another person in a manner causing sexual desire or shame, punishing both with the same penalty. Now the provision about recording someone to cause sexual desire or shame has

been applied even to a situation that does not involve nudity — for example, someone wearing leggings.[31] If the same standard is applied to this provision on editing, synthesizing, and fabricating images, synthesizing someone's face to a different body wearing leggings can also be punished harshly. The result will be even more unfair if the editing, synthesizing, and fabricating was done in such a nonrealistic manner that no reasonable person would believe the person featured had engaged in the presumably erotic situation, despite there being no nudity.

What is more dangerous is that even possession and viewing of such deepfake material is punishable by up to three years of imprisonment. This punishment violates one of the tenets of freedom of speech — that only the speech likely to cause "substantive" harm may be punished or otherwise restricted[32] — since possession and viewing of the existing material does not cause any harm to others.

In contrast, possession and viewing of child sexual abuse material (CSAM) is constitutionally punishable, but that is predicated on the theory that production of the material itself involves and victimizes real children and that the act of possession and viewing contributes to such abuse (i.e., production) by creating demand for its production. This theory makes sense because children are deemed legally incapable of consenting to sex, so any involvement of children in sexual activities during the production constitutes a crime. However, when the production itself does not involve such criminal activity, the theory does not apply. That is one reason the US Supreme Court has ruled that punishing nonrealistic material such as cartoons and animation as harshly as other CSAM (e.g., punishing possession of CSAM) is unconstitutional: The theory of generating demand for criminal activities does not apply to cartoons and animation.[33] Sexual deepfakes — though by definition involving real people — do not necessarily involve children, and therefore the visual consumption of their sexual activities does not in itself constitute a crime. The defamatory harm — an illusion that the victim is engaging in the depicted sexual activity — takes place only when such visual images are shared with a third party. Therefore, the South Korea law punishing possession of all sexual deepfakes requires a stronger justification.

2.5. Hate Speech

Only two laws in South Korea can be said to govern hate speech. One is the general German-style criminal insult law (Article 311 of the Criminal Code), which has often been used by the victims of hate speech but has been used much more vigorously by the professionals whose livelihoods critically depend on reputation, such as celebrities and politicians.[34] The other is a single provision in the Disability Discrimination Act (Article 32), which prohibits insulting comments against physically handicapped persons. There is no sign or precedent that AI-generated insults are to be more severely prosecuted or disciplined than other insults or hate speech.

31 Kim Na-young, "Man Fined in Retrial Over Illicit Filming of Woman in Leggings," Yonhap News Agency, November 2021, https://en.yna.co.kr/view/AEN20211103002100315.
32 Schenck v. United States, 249 U.S. 47 (1919).
33 Ashcroft v. Free Speech Coalition, 535 U.S. 234 (2002).
34 For instance, Kim Jae-heun, "K-Pop Stars Take Stern Actions Against Malicious Comments and False Accusations," *Korea Herald*, July 1, 2024, https://www.koreaherald.com/article/3426036; Open Net, "Prominent Politicians' Use of Criminal Insult Laws Are Deeply Troubling," August 29, 2019, https://www.opennetkorea.org/en/wp/2714.

2.6. Election and Political Content

As mentioned, AI-specific laws apply defamation-type norms more severely to AI-generated contents used in electoral contexts. Article 82-8 of the Public Officials Election Act (Election Campaigning Using Deepfake Video) reads:

> (1) No one shall produce, edit, distribute, exhibit or display the virtual sound, image or video made with artificial intelligence technology, etc., which is difficult to distinguish from the reality, within 90 days from the election day for the purpose of election campaign.
>
> (2) Anyone who produces, edits, distributes, exhibits or displays deepfake video outside the aforesaid period for the purpose of election campaign must label on the video to inform clearly its virtual character pursuant to the rules of the Central Election Commission. [Promulgated December 28, 2023]

The above provision prohibits the production and distribution of deepfakes for the purpose of an election campaign for 90 days before the election day. Significantly, there is no requirement that the deepfake mislead anyone about anything. A profile photo of a candidate, polished to make the subject appear younger or more passionate, will be the subject of criminal prosecution if such polishing was done through artificial intelligence "well (as if no polishing had been done)." Unlike similar state laws in the United States, no defense exists based on satire, parody, and hyperbole. Also, there is no requirement that the material interfere with the fairness of the related election.

2.7. Copyright

South Korea's Copyright Act currently does not recognize AI-generated content as eligible for copyright protection, given that authorship is limited to human creators. This stance aligns with international norms but raises questions about the legal status of AI-generated works. In December 2023, the Ministry of Culture, Sports and Tourism (MCST) reaffirmed the act's position, stating that AI-created content would not be granted copyright registration.[35] In line with this, the country's largest performing rights society, the Korean Music Copyright Association, requires all new song registrations to be backed by a written commitment that no AI was used in composing them.[36]

The lack of copyright protection for AI-generated content has implications for freedom of expression: It may encourage the use and dissemination of such content without any restriction on copying and multiplying the content since there are no intermeddling copyright holders.

The other side of this protection is whether copyrighted works can be used as the training data for AI. This is also related to freedom of speech as freedom of speech includes access to knowledge. Without AI, people have enhanced their knowledge by reading material on the internet directly, but more people are accessing knowledge through the summaries or paraphrasing done by AI, which reads the source material (and much more) for them. Whether the act of "reading" performed by an LLM is any different from the human act of "reading" is a crucial question that will decide the scope and quantity of the material upon which LLMs can be trained. In other words, overzealous enforcement of copyright protection on the training data

35 "Analysis of AI Regulatory Frameworks in South Korea," *Asian Business Law Journal*, April 15, 2024, https://law.asia/ai-regulatory-frameworks-south-korea/.
36 Yoon Min-sik, "Music Copyright Group Mandates 'No AI Use' for New Songs," *Korea Herald*, April 1, 2025, https://www.koreaherald.com/article/10455314.

against AI developers will reduce AI users' access to knowledge. For instance, if the *New York Times* and the *Washington Post* prohibit AI from scraping the facts from their news articles, the people depending on AI for obtaining knowledge from those sources, instead of reading *New York Times* and *Washington Post* articles directly, will receive an inferior set of knowledge.

On January 16, 2024, the MCST and the Korea Copyright Commission (KCC) released Guidelines on Generative AI and Copyright (the "Guidelines").[37]

AI service providers are encouraged to do the following:

- Secure legal basis for using any copyrighted works prior to using them given the current lack of clear legal standards on whether using copyrighted works for training AI models constitutes "fair use" under copyright law.

- Prevent copyright infringement by filtering out any expression that is identical or similar to copyrighted works from AI-generated outputs.

- Allocate liabilities among foundation model developers and downstream AI service providers who deploy such models in relevant contracts to help resolve future disputes that may arise from copyright infringement by AI-generated content.

- Invest in technologies and research to label AI-generated content with an ultimate goal to protect copyright holders' rights while also facilitating seamless use of copyrighted work.

Any copyright holders that do not want their copyrighted works to be used to train AI models are advised to clearly indicate such intent in relevant contracts or adopt technical measures to preclude such use by adding robot exclusion standards.

It is clear that the KCC is not supportive of the idea of freely allowing copyrighted works to be used for training AI under the "fair use" doctrine. More importantly, the Guidelines answers the question "Why is there a copyright issue in training AI?" by stating, without explanation, that use of copyrighted works in training AI "requires the consent of copyright holders" (p. 54).

However, since 2012, Korea has adopted the US-style "fair use" provision in preparation for the Korea-US Free Trade Agreement (signed in 2007),[38] which over time loosened the previously civil-law-ridden restrictive interpretation of fair use to a more liberal one.[39]

2.8. Measures Empowering Freedom of Expression

As noted, the new Lee Jae-myung administration promised in its campaign platform, "Everyone's AI (모두의 AI)," that high-grade AI would be available for all people in Korea for free. The native NPU project and the national data center project (à la the US Stargate), if successful, will contribute to such an initiative. However, it is too early to tell what specific outcomes will result from these efforts.

37 Kim & Chang, "Copyright in the Age of Artificial Intelligence," July 19, 2024, https://www.ip.kimchang.com/en/insights/detail.kc?idx=29913&sch_section=4.
38 Copyright Act, art. 35-2.
39 Nam Heesob, "Changes Induced by Open-Ended Fair Use Clause: Korean Experiences," InfoJustice, October 2016, https://infojustice.org/archives/37215.

2.9. Miscellaneous

2.9.1 Data Protection Law As Applied to Machine Learning Process

On top of the restrictive effect of data protection laws on the availability of personal data for machine learning, the machine learning itself is processing of personal data. Machine learning usually takes the form of data processing, which does not retain the personal data. For instance, AI will train itself on the health records of many individuals so it can later produce an answer to a prompt such as "what is the usual treatment for disease X?" without actually retaining the health records themselves. In doing so, the health records will first have been de-identified (i.e., anonymized, pseudonymized, or otherwise stripped off the identifying components of the data) and therefore brought out of the stricture of data protection laws and then "read" by the LLM system. This de-identification is crucial because otherwise the reading process would have necessitated the expensive and nearly impossible task of tracking down patients from years ago and asking whether their health records can be used for the new purpose unrelated to the original purpose of treatment: research.

A unique problem in South Korea is that the politicized debate between civil society and the industry/government ended up in the worst possible regulation, whereby de-identification (or equivalently pseudonymization), otherwise welcomed as a privacy-enhancing measure in other jurisdictions, became threatening and dangerous to the data subjects' rights.[40] In Korea, pseudonymization became a necessary condition for using the data for scientific research purposes, whereas GDPR — the original data protection law that the Korean law is modeled after — views pseudonymization as one of the important safety measures to be considered in authorizing nonconsensual use of the data. It may have been only good for data subjects' privacy that such safety measure was made an absolutely necessary condition for nonconsensual repurposing of their data. However, it went further: all pseudonymized data - even the ones pseudonymized for non-scientific purposes, were freed from data subjects' access rights or processing-halting rights. Given such sweeping power bestowed upon pseudonymization, the civil society in turn demanded and won a complete ban on re-identification (i.e., reattaching the personal identifiers to the de-identified data) in the country's data protection law (Article 28-5 of Personal Information Protection Act).

But such draconian provision boomeranged on the data subjects who wanted to exercise, for instance, access rights or processing-halting rights. When the data subjects actually inquired how and whether their data were used for certain scientific research or they wished to remove their data from such research, the data controllers simply answered that they cannot respond because of the absolute ban on re-identification of the data. Even if they could, once de-identified, the data are no longer under the strictures of data protection law. In response, the civil society are now demanding a moratorium on all pseudonymization since, under that measure, data subjects cannot object to or halt the processing of their data.

A positive development did take place: When Open Net demanded that only the data pseudonymized for research/archiving/statistical purposes are freed from data subjects' access and halt rights,[41] the provision was amended to reflect that (Article 28-7 of Personal Information Protection Act).

40 Natalie Pang and Kyung Sin Park, "Data Innovations and Challenges in South Korea from Legislative Innovations for Big Data to Battling COVID-19," in *Data and Innovation in Asia Pacific* (Konrad-Adenauer-Stiftung, 2021).
41 Open Net, "PIPA's misguided derogation on pseudonymized data puts privacy at risk", October 20, 2020, https://www.opennetkorea.org/en/wp/3127.

However, recently, the Supreme Court issued an unseemly decision that further confuses the discourse: *Pseudonymization, since it enhances privacy, does not constitute "processing"; therefore, it is not subject to the strictures of data protection law and especially to the data subjects' right to halt or object to specific data processing.*[42] A more palatable solution would be that pseudonymization be deemed "processing" but not be subject to the stringent consent requirement regarding data subjects because such a privacy-enhancing mode of processing is considered within the reasonable scope of the original purpose for which the data were collected. However, since it is processing, the data subjects are still entitled to the reasonable right to object to or access data processing. Also, the absolute ban on reidentification must be give way to a more flexible restriction where the data can be reidentified when data subjects wish to exercise their access or processing-halt rights.

42 Korean Supreme Court, 2025.7.18, Judgment 2024 Da 210554.

3. Conclusion

South Korea's AI Basic Act, modeled after the EU AI Act, imposes certain risk mitigation measures, both before and after creation and distribution, on certain *applications* of AI. To the extent that the regulated application does not present a unique risk, these measures can suppress the use of certain software in the decision-making process and therefore suppresses freedom of speech, where freedom of speech means freedom to speak through all mediums. As well, some risk mitigation measures are enforced by administrative bodies, whose intermeddling in the decision-making aspect of AI operation may constitute unjustified censorship. Both South Korean copyright law and data protection law suffer from uncertainty about how to characterize, respectively, machine reading and pseudonymization in the steps for training AI. The resolution of the former is likely to follow the "fair use" decisions from US courts, while clarifying pseudonymization in data protection will require a calm and non-polemicized discussion on how the GDPR has balanced data innovation and data protection around that concept. South Korea has specific laws that penalize the use of AI in electoral contexts and sexual contexts, both of which literally ban deepfakes. Finally, the nation's generally poor state of freedom of speech — with its criminal defamation laws, weak intermediary liability safe harbor, strong administrative censorship, and off-balance data protection provisions — will not only chill the efficient use of AI in concocting powerful speeches but also restrict the diversity and volume of data available for AI training.

THE FUTURE OF FREE SPEECH

OCTOBER 2025

www.futurefreespeech.org