



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.



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 dollars³, 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.cnn.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.

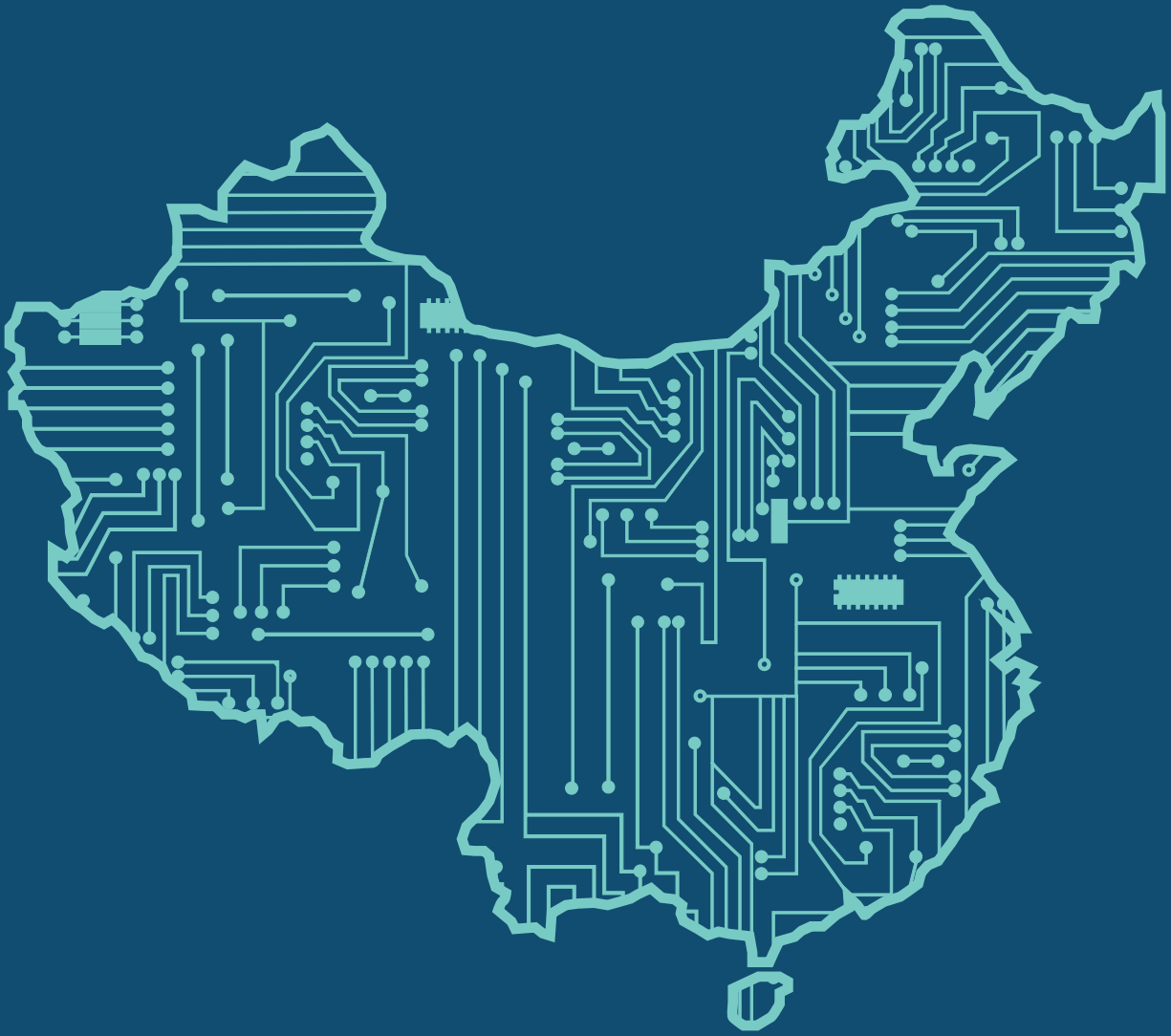
particular focus on elections and political speech, hate speech, defamation, explicit content (including child sexual abuse material and nonconsensual intimate images), and copyright. We also consider measures that actively promote freedom of expression, such as AI literacy initiatives and policies supporting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

Ge Chen*

* Associate professor of Global Media and Information Law and director of the Centre for Chinese Law and Policy, Durham Law School; affiliated fellow, Information Society Project, Yale Law School; associate, Centre for Intellectual Property and Information Law, University of Cambridge.

Abstract

This chapter examines China’s regulatory approach to generative AI and its implications for freedom of expression. While China has not enacted a stand-alone law on AI, it has developed a comprehensive patchwork of administrative measures and platform obligations that form a de facto regime of control over AI-generated content. Key frameworks include cybersecurity laws, administrative rules governing generative AI services, and interrelated rules targeting deep synthesis technologies and algorithmic recommendation systems.

These frameworks collectively impose ideological, political, and technical constraints on generative AI systems, requiring alignment with socialist core values, national security objectives, and censorship norms. Platforms and developers are obliged to prescreen, filter, and monitor AI outputs, including for the purpose of combating copyright infringement, defamation, obscenity, misinformation, and any content deemed politically sensitive. Importantly, vague legal concepts such as “public order” and “social morality” are interpreted expansively to enable broad discretionary enforcement.

China treats generative AI as a uniquely high-risk technology and imposes heightened liability and oversight regimes that significantly constrain its expressive potential. Although no formal distinction is made between human- and AI-generated content in law, AI outputs are often subjected to greater scrutiny due to their perceived uncontrollability and scale. Regulations also target specific use cases — such as AI-generated political commentary, sexually suggestive content, or satirical speech — under the guise of maintaining ethical standards or information security.

Overall, China’s regulatory model emphasizes anticipatory censorship and political control, offering little room for protective measures like liability exemptions, independent oversight, or rights-based challenges. Rather than empowering users or protecting freedom of expression, the governance of generative AI in China reinforces existing authoritarian speech controls and extends them into emerging technological domains.



Ge Chen

Ge Chen is an associate professor of Global Media and Information Law and director of the Centre for Chinese Law and Policy, Durham Law School; affiliated fellow, Information Society Project, Yale Law School; associate, Centre for Intellectual Property and Information Law, University of Cambridge. He is the recipient of the 2025 Franklyn S. Haiman Award for Distinguished Scholarship in Freedom of Expression, awarded by the U.S. National Communication Association.

1. Introduction

Traditionally, China's approach to freedom of expression has been markedly restrictive, reflecting the political ethos of an authoritarian state dominated by the Chinese Communist Party (CCP). Although Article 35 of China's Constitution nominally recognizes freedom of speech, decades of political censorship and ideological control have overshadowed this constitutional promise.¹ The recent decade under Xi Jinping has seen a significant intensification of censorship domestically, paired with increasingly assertive international censorship and propaganda strategies that may well influence freedom of expression in other parts of the world.² Amendments to the People's Republic of China (PRC) Constitution in 2018 abolished presidential term limits and entrenched the CCP's supreme constitutional status by codifying "overall party leadership" as a defining constitutional principle,³ enhancing the CCP's institutional control over all aspects of governance, including speech regulation.

In this context, artificial intelligence (AI) emerges as both a pivotal tool and a significant regulatory challenge within China's broader constitutional and governance framework. Recognizing the profound implications of generative AI technologies like ChatGPT, Chinese authorities have swiftly moved to legislate their development and deployment. Regulations promulgated by bodies such as the Cyberspace Administration of China (CAC) explicitly define AI governance in accordance with "socialist core values,"⁴ demanding adherence to national security considerations. Recent legislative initiatives, including the joint regulatory efforts of central government departments targeting AI-generated content since 2021,⁵ reflect a comprehensive attempt to embed AI regulation within a political framework explicitly prioritizing state power, ideological conformity, and social stability.⁶ These laws specifically prohibit AI-generated content from "subverting state power," "inciting secession," or "disrupting economic and social order,"⁷ thus embedding traditional CCP ideological controls into the cutting-edge domain of generative AI.

This intersection of AI technology with China's constitutional and political imperatives underscores the evolving complexity and global significance of China's regulatory strategies, posing critical implications for international norms around freedom of expression and the governance of emerging technologies. The following section analyzes the specific implications of China's AI law and policies from different perspectives that are most relevant to the exercise of freedom of expression.

1 For a detailed account, see Ge Chen, "The Constitutional Rise of Chinese Speech Imperialism," *Journal of Free Speech Law* 2, no. 2 (2023): 501-15, <https://www.journaloffreespeechlaw.org/chen.pdf>.

2 Chen, "Chinese Speech Imperialism," 558-63.

3 PRC Const. art. 1 (2018).

4 Interim Measures for the Management of Generative AI Services (hereafter cited as IMMGAIS) [生成式人工智能服务管理暂行办法], art. 4(1), August 15, 2023.

5 See, e.g., Guiding Opinions on Strengthening Comprehensive Governance of Internet Information Service Algorithms (hereafter cited as Guiding Opinions) [关于加强互联网信息服务算法综合治理的指导意见], September 17, 2021; Provisions on the Administration of Algorithm Recommendations for Internet Information Services (hereafter cited as PAARIIS) [互联网信息服务算法推荐管理规定], March 1, 2022; Provisions on the Management of Deep Synthesis of Internet Information Services (hereafter cited as PMDSIIS) [互联网信息服务深度合成管理规定], January 10, 2023; AI Security Governance Framework [人工智能安全治理框架], September 9, 2024; Administrative Measures on Labeling AI-Generated Synthetic Content (hereafter cited as AMLAIGSO) [人工智能生成合成内容标识办法], March 7, 2025.

6 In the context of AI, this is condensed in the State Council's administrative regulations concerning internet content regulation. Internet Information Services Management Measures (hereafter cited as IISMM) [互联网信息服务管理办法], art. 15, November 22, 2024.

7 IMMGAIS, art. 4(1); PAARIIS, art. 6.

2. Substantive Analyses

2.1. General Standards of Freedom of Expression

Under the PRC Constitution, speech regulation has been deeply intertwined with CCP ideology, systematically suppressing dissenting political speech and prioritizing “national security” and “public order.”⁸ The 2018 constitutional amendments, which further solidified CCP leadership, empower CCP organs significantly, positioning intra-party rules and disciplinary actions above constitutional norms on free speech.⁹ Consequently, constitutional protections for free speech remain largely symbolic, as enforcement prioritizes party-defined restrictions and ideological conformity.

Most importantly, Chinese courts function merely as political instruments of the CCP, entirely subordinated to the party’s political leadership and completely deprived of authority to apply constitutional provisions in judicial practice.¹⁰ Consequently, constitutional rights remain theoretical rather than practical. No court judgment in the PRC ever cites, interprets, or evaluates the right to freedom of speech under Article 35 or other relevant constitutional provisions against competing legal interests, a practice common in the European Court of Human Rights or the US Supreme Court when adjudicating cases involving this right.

Internationally, China has signed but not ratified the International Covenant on Civil and Political Rights, signaling formal acknowledgment without substantive commitment to international free speech norms.¹¹ Domestically, CCP dominance ensures that free speech perceived as critical of the party or leadership is swiftly curtailed under sweeping national security laws, such as the 2015 National Security Law and the 2017 Cybersecurity Law, both of which broadly criminalize speech that the Chinese government deemed could potentially disrupt the CCP’s monopoly of power, subvert or incite the subversion of state power, overthrow of the socialist system, incite the secession of the country, or undermine national unity.¹²

Illustrative cases, including recent crackdowns on activists and journalists reporting sensitive issues such as the COVID-19 responses and the Russia-Ukraine War,¹³ exemplify how constitutional amendments prioritizing the CCP’s leadership effectively nullify constitutional guarantees of free expression at home. Internationally, these amendments underpin China’s expanding transnational censorship, extending its speech regulation through physical threatening, digital platforms, and economic pressures to suppress criticism and control global narratives about China,¹⁴ significantly challenging international free speech standards.

8 PRC Const. arts. 53, 54; Chen, “Chinese Speech Imperialism,” 517–25.

9 Chen, “Chinese Speech Imperialism,” 527–39.

10 Ge Chen, “Piercing the Veil of State Sovereignty: How China’s Censorship Regime into Fragmented International Law Can Lead to a Butterfly Effect,” *Global Constitutionalism* 3, no. 1 (2014): 38, <https://doi.org/10.1017/S2045381713000282>.

11 Chen, “Piercing the Veil,” 45–50.

12 National Security Law of the PRC, art. 15, July 1, 2015; Cybersecurity Law of the PRC, art. 12, June 1, 2017.

13 Kieran Green et al., *Censorship Practice of the People’s Republic of China* (Center for Intelligence Research, February 20, 2024), 34–40.

14 The Congressional-Executive Commission on China, 118th Cong., *Annual Report 2024* (2024), 290–95.

2.2. AI-Specific Legislation and Policies

The rapid expansion of AI technologies in China has triggered substantial legislative developments aimed at managing the significant societal impacts and associated risks. China’s regulatory approach to AI is characterized by detailed and multilayered rules designed explicitly to ensure that AI development conforms to CCP-defined “socialist core values”¹⁵ and stringent “national security” criteria.¹⁶

2.2.1. The Basic Structure

Central to this regulatory framework are the Cybersecurity Law, the Data Security Law (DSL),¹⁷ and the Personal Information Protection Law (PIPL)¹⁸ — foundational statutes governing the collection, storage, processing, and use of data by AI systems. These laws serve as critical pillars for speech regulation in the AI era, establishing stringent requirements for cybersecurity, data governance, and personal data protection. Their impact on AI firms is particularly significant, as these regulations shape not only the technical handling of data but also the broader mechanisms of content moderation and speech control under the CCP’s oversight.¹⁹

Unlike the European Union’s comprehensive, principles-based governance framework, China adopts a more fragmented, application-specific approach to regulating AI. Specific departmental regulations target distinct technological applications,²⁰ data governance,²¹ and content regulation,²² as well as their perceived impacts on societal ethics.²³ This partitioned approach enables more stringent speech control and flexibility for the state apparatus. Key recent AI regulatory actions with specific implications for free speech include the Interim Measures for the Management of Generative AI Services (IMMGAS), which is currently China’s primary regulation governing generative AI. This comprehensive regulation mandates careful and risk-based oversight of AI services, explicitly banning AI-generated content deemed to contravene existing rules of content regulation covering political speech, copyright, defamation, explicit content, and the like.²⁴

2.2.2. Content Regulation

Rather than merely applying general content standards to new technologies, the IMMGAS codify a unique and binding obligation to use AI in service of state-defined ideological conformity — making this regulation a direct legal mechanism for restricting speech through AI far beyond the baseline limitations typically seen in liberal democracies. Chinese generative AI models are thus subject to a highly restrictive regulatory framework that imposes broad obligations on the provenance and processing of training data. All AI developers and service providers are obliged to incorporate these content controls into all phases of model development, including data selection, algorithm design, and service delivery. Developers must ensure, for example, that data used for training not only is legally sourced but also adheres to ideological, ethical, and technical standards — ranging from avoiding discriminatory biases to ensuring “authenticity” and alignment with socialist core values.²⁵

¹⁵ For instance, the Chinese government issued the Guiding Opinions on Algorithm Governance in 2021, which states that the goal of China’s algorithm governance is to establish a comprehensive governance pattern of algorithm security with “correct orientation of algorithms,” “core socialist values,” and “correct political direction, public opinion orientation, and value orientation in the application of algorithms” to disseminate “positive energy.” Guiding Opinions, part IV(12).

¹⁶ PAARIS, art. 6.

¹⁷ The Data Security Law of the PRC (DSL), June 10, 2021.

¹⁸ The Personal Information Protection Law of the PRC (PIPL), August 20, 2021.

¹⁹ Chen, “Chinese Speech Imperialism,” 557–58.

²⁰ See, e.g., PAARIS and PMDSIS.

²¹ DSL and PIPL.

²² Cybersecurity Law and ISMM.

²³ Measures for the Review of Science and Technology Ethics (Trial Implementation) [科技伦理审查办法(试行)], September 7, 2023.

²⁴ IMMGAS, art. 4(1)–(4).

²⁵ IMMGAS, art. 4(1)–(4).

These requirements go beyond basic technical safeguards and extend into content and viewpoint regulation, effectively filtering out politically or socially sensitive material at the foundational level of model development. In practice, this means that training data that may be politically sensitive or ideologically nonconforming must be excluded at the outset. Similarly, model outputs are expected to avoid content deemed “harmful,” “untrue,” or contrary to state-defined norms.²⁶ In effect, the design and training process is shaped by political and ideological filtering, structurally limiting the expressive potential of generative AI. These systemic content-based restrictions function as a form of preemptive censorship, significantly curtailing freedom of expression through technical design mandates.

2.2.3. Restriction of Specific Generative AI Models

China imposes a de facto ban on access to most foreign-developed generative AI models, including those from OpenAI such as ChatGPT.²⁷ While these restrictions are not always framed as explicit model-specific prohibitions, the combined effect of content regulation, cybersecurity oversight, and ideological control effectively prevents the deployment of foreign AI systems that do not conform to Chinese governance standards.²⁸

2.2.4. Open-Source Limitation

Contrary to some international frameworks — such as the EU’s AI Act, which offers exemptions for certain open-source AI models — China’s current AI regulatory system does not distinguish between open-source and proprietary models. Governance obligations are determined by the function, deployment context, and perceived risk of the AI system — particularly whether it has public opinion attributes or social mobilization potential. Although a 2024 expert draft of an AI Model Law proposed reduced liability for open-source providers who implement sufficient safety measures,²⁹ this remains a theoretical suggestion and has not been codified.³⁰

2.2.5. Risk Classification and Mandatory Labeling

The classification of generative AI as a politically sensitive and potentially destabilizing technology serves as both a justification for its determination as a high risk and a tool for intensified speech control. The IMMGAIS articulate a governance philosophy that tightly links AI development to national security and ideological control as the regulation target,³¹ while affirming a dual emphasis on “development and security” alongside “classified and tiered supervision.”³² In practice, this translates into significantly higher regulatory burdens for generative AI providers, creating a systemic chilling effect on expressive diversity and innovation.

China has enacted the Administrative Measures on Labeling AI-Generated Synthetic Content (AMLAIGSC), effective from September 1, 2025, which establishes a comprehensive mandatory labeling regime. Service providers must add explicit labels — such as text, audio, or visual indicators — on AI-generated content

²⁶ IMMGAIS, art. 4(1)–(4).

²⁷ Although Chinese companies may sometimes train domestic models using components of foreign systems, these models are subject to strict content controls and do not provide unfiltered access to the original tools. As a result, users in China are largely cut off from the expressive potential and informational diversity that global generative AI models offer. See US Select Committee on the CCP, *DeepSeek Unmasked: Exposing the CCP’s Latest Tool for Spying, Stealing, and Subverting U.S. Export Control Restrictions* (April 16, 2025), 6–9, <https://selectcommitteeontheccp.house.gov/sites/evo-subsites/selectcommitteeontheccp.house.gov/files/evo-media-document/DeepSeek%20Final.pdf>.

²⁸ See IMMGAIS, arts. 2, 20, 21.

²⁹ Artificial Intelligence Law of the PRC (Draft for Suggestions from Scholars), art. 90, April 26, 2024.

³⁰ Johanna Costigan, “China’s New Draft AI Law Prioritizes Industry Development,” *Forbes*, March 24, 2024, <https://www.forbes.com/sites/johannacostigan/2024/03/22/chinas-new-draft-ai-law-prioritizes-industry-development/>.

³¹ See IMMGAIS, art. 1.

³² IMMGAIS, art. 3.

including text, images, audio, video, and virtual scenes.³³ They must also embed implicit (metadata-based) labels for traceability. Platforms disseminating such content must verify and display these labels and even apply presumptive labeling where metadata is missing but content shows AI characteristics.³⁴ These requirements apply broadly and are enforceable through multi-agency oversight. The obligations, while intended to enhance transparency, impose stringent formalities and monitoring responsibilities that may impact anonymous or artistic expression.

2.2.6. Platform Duty for Almost Real-Time Removal of AI Content

Chinese law imposes proactive, sweeping, and urgent duties on platforms to detect and remove sensitive AI-generated content. AI service providers must promptly stop generation and transmission of and eliminate unlawful content once detected, and then report it to authorities.³⁵ They are also required to retrain models to prevent recurrence and take enforcement action (such as account suspension) against violators. Similarly, the law on algorithm recommendation requires immediate cessation of transmission and removal of illegal content, with mandatory reporting obligations.³⁶ These AI-specific rules exist alongside a broader, entrenched censorship framework mandating near real-time takedown of politically or socially sensitive information. While no precise time limit (e.g., 24 or 48 hours) is codified, the cumulative regulatory language — such as “immediately” and “timely” — often imposes a quasi-real-time removal requirement, particularly for politically sensitive content.

Overall, China’s evolving regulatory landscape for AI underscores a deliberate strategy to integrate technological governance with stringent ideological controls through speech regulation, reflecting broader efforts to sustain CCP authority while positioning China as a global leader in AI development.

2.3. Defamation

Traditionally, China’s defamation law framework, rooted in constitutional and statutory provisions, significantly constrains freedom of expression. Article 38 of the PRC Constitution explicitly protects personal dignity, prohibiting insults, libel, false accusations, or incrimination. Although constitutionally framed as protecting individuals against reputational harm, China’s defamation and libel law in the AI era functions broadly as a tool to suppress speech critical of individuals or entities closely associated with CCP interests.

2.3.1. Censorship Under Traditional Rules of Defamation and Libel

In China, defamation concerning private individuals is primarily regulated by private law. While these rules about reputational harm do not explicitly reference freedom of speech, they implicitly provide some guidelines for balancing reputation rights against expressive interests. In principle, online speech that insults or defames others can constitute a tort under China’s Civil Code, which explicitly prohibits damaging another person’s reputation through insults or defamatory statements.³⁷ The law provides that media reporting or public oversight conducted in the public interest does not constitute an infringement of reputation rights. Exceptions include cases where the report fabricates or distorts facts, fails to exercise reasonable diligence in verifying seriously inaccurate information provided by others, or uses insulting language to damage a person’s

³³ AMLAIGSC, art. 4.

³⁴ AMLAIGSC, arts. 5–6.

³⁵ See IMMGAIS, art. 14.

³⁶ See PAARIIS, art. 9.

³⁷ The Civil Code of the PRC, art. 1024.

reputation.³⁸ Consequently, any AI-generated content containing such allegedly defamatory speech could be considered tortious under Chinese law. Additionally, if an individual can substantiate that the content of media reports or online posts is false and harms their reputation, they are entitled to request corrective actions, including corrections or deletion.³⁹

However, under China's criminal defamation framework, online expression is subject to markedly stricter censorship than traditional forms of speech, wherein civil defamation cases can escalate into criminal offenses if deemed "serious."⁴⁰ According to judicial interpretations, defamatory online content viewed over 5,000 times or shared more than 500 times qualifies as "serious defamation,"⁴¹ directly tying criminal liability to the scale of digital dissemination. This quantitative standard places a substantial burden on content creators, compounded further by provisions criminalizing statements that lead to severe psychological or physical consequences, including suicide or self-harm.⁴²

2.3.2. Digital Defamation Rules Used as a Tool for Censorship

Importantly, these judicial interpretations also broadly criminalize the use of information networks to "defame others" if that defamation is deemed to "seriously endanger social order and national interests."⁴³ This approach, in fact, employs intentionally vague and expansive language that creates potential liability for "seditious libel": In recent years, numerous cases have involved individuals being prosecuted, tried, and sentenced on charges of defamation and slander specifically for criticizing Mao Zedong, Xi Jinping, other party-state leaders, or the CCP.⁴⁴

Under these principles, China has expanded its defamation framework through legislation specifically targeting criticism or negative portrayals of individuals officially designated as "heroes" or "martyrs." A 2018 law prohibits distorting, defaming, or denying the achievements and spirit of these officially recognized figures, explicitly extending these prohibitions to online media.⁴⁵ Similarly, the CAC expressly prohibits internet content providers from disseminating material that "harms national honor and interests," "distorts, vilifies, defiles, or denies the deeds and spirit of heroes and martyrs," or "sensationalizes gossip, scandals, and misdeeds."⁴⁶ Given generative AI's capacity to rapidly produce satirical or critical commentary, such broad restrictions pose acute risks to AI-generated content involving sensitive historical and political figures, necessitating stringent content moderation practices by AI service providers to avoid legal repercussions.

Several illustrative cases exemplify the practical enforcement of these regulatory frameworks. For example, in 2018, an individual faced penalties for posting allegedly derogatory online comments about a deceased firefighter, demonstrating the stringent protection afforded to public servants' reputations.⁴⁷ Similarly, a judicial decision against a blog post questioning the government's official account of the deaths of five soldiers during World War II reflects how historical narratives of "heroes and martyrs" are strictly controlled.⁴⁸ Specialized Internet Courts established in

³⁸ Civil Code, art. 1025.

³⁹ Civil Code, art. 1028.

⁴⁰ Criminal Law of the PRC, art. 246(1), December 29, 2023.

⁴¹ Interpretation of the Supreme People's Court and the Supreme People's Procuratorate on Several Issues Concerning the Application of Law in Handling Criminal Cases of Defamation and Other Crimes Committed via Information Networks [最高人民法院、最高人民检察院关于办理利用信息网络实施诽谤等刑事案件适用法律若干问题的解释], art. 2(1), September 6, 2013.

⁴² Interpretation of the Supreme People's Court, art. 2(2).

⁴³ Interpretation of the Supreme People's Court, art. 3.

⁴⁴ See Chen, "Chinese Speech Imperialism," 512-13.

⁴⁵ Law of the PRC on the Protection of Heroes and Martyrs, art. 22, April 27, 2018.

⁴⁶ Provisions on Ecological Governance of Network Information Content (hereafter cited as PEGNIC) [网络信息内容生态治理规定], arts. 6(3), 6(4), and 7(2), December 15, 2019.

⁴⁷ Xu Chang — Defaming a Martyred Firefighter, Intermediate People's Court of Yantai, Shandong, Civil Judgment, (2018) Lu 06 Civil First Instance no. 211 (June 26, 2018).

⁴⁸ Kiki Zhao, "Chinese Court Orders Apology Over Challenge to Tale of Wartime Heroes," *New York Times*, June 28, 2016, <https://www.nytimes.com/2016/06/29/world/asia/china-hong-zhenkuai-five-heroes.html>.

cities such as Beijing and Hangzhou increasingly adjudicate online reputational harm disputes,⁴⁹ institutionalizing speech regulation within digital spaces. These courts frequently address cases involving AI-generated or AI-assisted online content, reinforcing legal norms favoring reputational protection over free speech.

2.3.3. Stricter Liability for AI-Generated Defamatory Content

Recent AI-specific regulations explicitly integrate these long-standing defamation principles into the governance of generative AI, mandating that generative AI services “respect the legitimate rights and interests of others,” prohibiting content that “harms others’ physical and mental health” or infringes upon portrait rights, reputation rights, honor, privacy, and personal information.⁵⁰ By applying traditional defamation standards to AI-generated content, these new regulations impose heightened content moderation obligations on AI service and content providers. For example, the aforementioned quantitative criteria defining criminal liability under defamation law are particularly impactful for users of generative AI, whose outputs can rapidly achieve widespread dissemination.

Other recent regulatory developments further illustrate China’s stringent approach to managing potentially defamatory AI-generated speech, particularly through rules targeting “deep synthesis” (deepfake) technologies. One such regulation explicitly prohibits using deep synthesis technologies to create or disseminate content that endangers national security, damages national image, or infringes upon individuals’ lawful rights and interests.⁵¹ Specifically, these regulations mandate that providers conduct formal security assessments if they offer AI services and content involving biometrics (such as synthesized faces or voices) or sensitive content that could implicate national security or public interests.⁵²

Further, these regulations require content and service providers to conspicuously label AI-generated or AI-edited content to alert users of its artificial origins, particularly where the public might otherwise be confused or misled.⁵³ This obligation covers various services, including AI-generated dialogues (chatbots or intelligent writing services), voice synthesis technologies capable of mimicking or significantly altering individuals’ vocal characteristics, face replacement or manipulation technologies (such as deepfake videos altering public figures’ statements or actions), and immersive virtual scenarios.⁵⁴

In practice, this means that both individual users and platforms using generative AI to produce lifelike depictions of public officials, celebrities, or historical figures must clearly label such content as artificially generated. Likewise, deepfake videos or audio clips altering politicians’ statements or appearances must undergo rigorous labeling and moderation to prevent public confusion or manipulation. Such stringent requirements under the aegis of prohibiting defamatory or libelous information reflect government authorities’ anxieties about AI technologies potentially creating critical voices, deepening societal polarization, or destabilizing established political narratives. As a result, users employing AI to generate and disseminate content may unwittingly incur criminal liability, placing substantial pressure on AI developers and platforms to implement proactive filtering and moderation mechanisms to censor materials that “hype up gossips and scandals” and “promote indecency, vulgarity, and kitsch.”⁵⁵

49 Yuan Yuan, “Plugged In: Internet Courts Make It Easier to Access Judiciary Procedures,” *Beijing Review*, January 9, 2020, https://www.bjreview.com/China/202001/t20200109_800189546.html.

50 IMMGAIS, art. 4(4).

51 PMDSIIS, art. 6.

52 PMDSIIS, art. 15.

53 PMDSIIS, art. 17.

54 PMDSIIS, art. 17.

55 PEGNIC, art. 7(2) and 7(7).

In sum, China's defamation and libel laws, now deeply embedded within the evolving regulatory landscape for AI governance, significantly restrict freedom of expression. AI-generated content faces intense scrutiny under expansive definitions of reputational harm and stringent criminal liability thresholds, reinforcing broader CCP political objectives of speech control and ideological conformity.

2.4. Explicit Content

2.4.1. General Censorship of Sexually Explicit Content

In China, all sexually explicit content, including child sexual abuse material (CSAM) and non-consensual intimate imagery (NCII), has historically been regulated under broad and severe censorship frameworks. The Chinese legal approach toward sexually explicit materials primarily emphasizes moral standards and social stability. Under the Criminal Code, the production, dissemination, and sale of pornography, particularly CSAM, are explicitly prohibited, carrying severe criminal penalties.⁵⁶ The law defines obscenity ambiguously, enabling wide discretionary enforcement.⁵⁷ This broad categorization effectively allows authorities to suppress any content they perceive as detrimental to public morality or social order, resulting in suppression frequently extending into areas of personal sexual autonomy and non-traditional sexual orientations.⁵⁸

Furthermore, NCII (commonly known as “revenge porn”) is increasingly regulated under provisions concerning privacy and personal dignity.⁵⁹ While the legislative intent ostensibly targets protection of individual rights, in practice, enforcement of these laws often overlaps with broader censorship aimed at controlling the dissemination of all sexually explicit material — irrespective of consent — thereby reinforcing the state's moral oversight.⁶⁰

The Chinese government also employs administrative mechanisms to strengthen this regulatory regime. For instance, the CAC explicitly prohibits producing or disseminating any online content classified as “obscene or pornographic.”⁶¹ Under these regulations, content producers and online platforms must proactively monitor, censor, and report potentially explicit materials, strengthening the government's strict control over sexual expression in digital spaces.⁶²

2.4.2. Censorship of Politically Sensitive Explicit Content

Historically and contemporarily, sexually explicit material in China often intersects with political sensitivity. The term “pornography” often encompasses sexually explicit content deployed as political critique or satire directly challenging state authority and ideological norms. Such content has been particularly sensitive for Chinese authorities that systematically deploy censorship mechanisms under the guise of moral regulation to suppress politically dissenting expressions.⁶³

⁵⁶ Criminal Law, arts. 363–65.

⁵⁷ Criminal Law, art. 367.

⁵⁸ See, e.g., Graeme Reid, “China's Pornography Laws Are a Backdoor for Censorship,” *Human Rights Watch*, November 29, 2018, <https://www.hrw.org/news/2018/11/29/chinas-pornography-laws-are-backdoor-censorship>.

⁵⁹ Ding Guofeng and Luo Shasha [丁国锋 罗莎莎], “Posting a Private Video Showing Someone's Face to a Porn Site in Retaliation Was Found Guilty of Insults [为报复将他人露脸隐私视频发黄网被认定犯侮辱罪]” *CCTV.com*, February 5, 2024, <https://news.cctv.com/2024/02/05/ARTIj547VmrByZguyXX6Qik2240205.shtml>.

⁶⁰ Chu Cheng, “Incomplete and Opaque: The Problems with China's Porn Laws,” *Sixth Tone*, December 9, 2016, <https://www.sixthtone.com/news/1661>.

⁶¹ PEGNIC, art. 6(9).

⁶² IISMM, art. 15(7).

⁶³ Y. Yvon Wang, “Yellow Books in Red China: A Preliminary Examination of Sex in Print in the Early People's Republic,” *Twentieth-Century China* 44, no. 1 (2019): 85–87, <https://muse.jhu.edu/article/713397>.

China's approach to dealing with pornographic materials fundamentally differs from liberal democratic traditions, where sexual and political speech are treated as separate legal categories. In practice, China employs a comprehensive censorship model led by the National Office Against Pornographic and Illegal Publications — another party-state organ directly responsible for the Propaganda Department of the Central Committee of the CCP — blending moral governance with political control through regular crackdown campaigns.⁶⁴ This approach allows authorities to continually produce new legal justifications for censorship, particularly targeting sexual expression with implicit or explicit political commentary.⁶⁵ The state's vague definitions and arbitrary enforcement practices — combined with public tolerance for suppressing sexually explicit material — enable authorities to seamlessly merge political repression with moral oversight, stifling politically charged sexual speech without overt backlash.⁶⁶

2.4.3. AI-Specific Censorship of Explicit Content

Recently, China has extended its established censorship regime explicitly into the realm of AI-generated content. The regulations specifically governing generative AI impose comprehensive obligations on service providers to adhere to broader ideological and moral standards dictated by the state.

Notably, the IMMGAIS prohibit AI-generated content that promotes or disseminates “violence, obscenity, and pornography,” as well as political subversion or threats to national unity.⁶⁷ These regulations reflect the overall political and moral stance of the Chinese government, embedding traditional norms of pornography censorship directly into cutting-edge generative technologies.⁶⁸ Consequently, platforms hosting AI-generated content are obliged to implement proactive filtering and moderation mechanisms to censor materials containing “sexual innuendo or provocations that easily cause sexual fantasies.”⁶⁹

Furthermore, the recent regulations on deepfake technologies represent another critical regulatory advancement. These provisions mandate that AI-generated deepfake content — especially images, audio, or video that significantly alter personal identity features or simulate realistic scenarios — must be clearly marked to prevent public confusion or misuse.⁷⁰ This rule applies to various generative AI services such as intelligent dialogues, human voice synthesis, face-swapping, and immersive realistic scenarios.⁷¹ Moreover, providers of these services and content are required to undertake stringent security assessments, particularly when generated content involves biometric data (faces, voices) or politically sensitive contexts.⁷²

Such deepfake technologies are capable of generating explicit imagery or videos that could easily become politically weaponized if they depict public figures, which is why they have incited official concern. Consequently, Chinese authorities require extensive labeling, preemptive moderation, and strict content control mechanisms to preclude the emergence and dissemination of politically sensitive or morally contentious deepfakes.⁷³ Platforms providing these services must proactively screen and label potentially sensitive AI-generated content, effectively implementing a robust system of self-censorship consistent with state directives.

64 US Department of State, *China (Includes Tibet, Hong Kong, and Macau) 2019 Human Rights Report* (2019), 30.

65 See Mei Ning Yan, “Regulating Online Pornography in Mainland China and Hong Kong,” in *Routledge Handbook of Sexuality Studies in East Asia*, eds. Mark McLelland and Vera Mackie (Routledge, 2014), 388–89.

66 See, e.g., Mascha Borak, “China's Porn Censors Shut Down 12,000 Websites in the First Half of 2020,” *South China Morning Post*, July 9, 2020, <https://www.scmp.com/abacus/news-bites/article/3092512/chinas-porn-censors-shut-down-12000-websites-first-half-2020>.

67 IMMGAIS, art. 4(1).

68 Katie Wickens, “China's ‘Mind-Reading’ Porn Detection Cap Takes Censorship to New Levels,” *PC Gamer*, July 15, 2022, <https://www.pcgamer.com/chinas-mind-reading-porn-detection-cap-takes-censorship-to-new-levels/>.

69 PEGNIC, art. 7(4).

70 PMDSIIS, art. 17.

71 PMDSIIS, art. 17(1)–(4).

72 PMDSIIS, art. 15.

73 AMLAIGSC, arts. 4–6.

Additionally, China's Regulations on Algorithmic Recommendation Services establish specific obligations regarding minors. Platforms are expressly forbidden from algorithmically recommending content potentially harmful to minors, including sexually explicit materials or content that could incite unsafe behaviors or violate social morality.⁷⁴ This creates a two-tiered obligation: Not only must AI-generated explicit content be rigorously monitored and filtered, but algorithmic recommendations must also be meticulously adjusted to protect minors, further expanding the censorship and compliance obligations placed on AI service providers.

In sum, the introduction of AI-specific regulations integrates seamlessly into China's established framework of moral and political censorship. Providers of generative AI must navigate a complex legal environment requiring continual content filtering, rigorous security assessments, and proactive moderation to avoid serious legal repercussions.

2.5. Hate Speech

The legal framework governing hate speech in China is intrinsically embedded within the broader strategy of speech regulation, emphasizing state interests and social stability over individual rights. Although formally rooted in principles of equality, this approach frequently subordinates genuine anti-discrimination protections to the overarching political objectives of regime stability and ideological conformity.

2.5.1. Chinese Hate Speech Laws as a Tool for Censorship

China's constitution explicitly affirms equality among ethnic groups, prohibits discrimination, and guarantees equal protection to all citizens.⁷⁵ Despite these nominal protections, hate speech laws predominantly serve political rather than genuine anti-discrimination purposes. The foundational principle underlying Chinese hate speech regulation prioritizes ideological and political uniformity, national security, and unity over substantive equality.⁷⁶ Thus, equality rights articulated within the constitutional text are operationalized primarily within a context that underscores political stability, state control, and regime legitimacy.

China's Criminal Code addresses hate speech by criminalizing severe instances of inciting ethnic hatred or discrimination.⁷⁷ However, courts apply vague thresholds in assessing hate speech offenses, linking criminal liability directly to broad social impacts rather than clearly defined individual harms.⁷⁸ This vagueness creates significant uncertainty, exacerbating the risks for speakers who might unwittingly push against these amorphous boundaries. Additionally, administrative regulations reinforce this stringent approach, mandating preemptive moderation of online content and penalizing internet service providers (ISPs) that fail to effectively censor hate speech or politically sensitive content.⁷⁹ Consequently, China's legal environment fosters extensive proactive self-censorship among platforms.

2.5.2. Censoring Online Hate Speech in Legal Practice

Under this regulatory framework, administrative laws governing online speech require ISPs to actively filter content that could be considered hate speech. For instance, core administrative regulations such

⁷⁴ PAARIS, art. 18.

⁷⁵ PRC Const. arts. 4, 33, 36.

⁷⁶ Ge Chen, "How Equalitarian Regulation of Online Hate Speech Turns Authoritarian: A Chinese Perspective," *Journal of Media Law* 14, no. 1 (2022): 170-71, <https://doi.org/10.1080/17577632.2022.2085013>.

⁷⁷ Criminal Law, arts. 249 and 250.

⁷⁸ Chen, "Online Hate Speech," 173-74.

⁷⁹ Chen, "Online Hate Speech," 171-73.

as the Internet Information Services Management Measures (IISMM) and the Provisions on Ecological Governance of Network Information Content (PEGNIC) broadly prohibit content that “incites ethnic hatred or discrimination,” “damages national unity,” or “contravenes state religious policies.”⁸⁰

In practice, however, these provisions are enforced selectively and instrumentally, upholding existing biases within China’s censorship and speech-control architecture. Since the overarching priority for internet regulators is regime stability, conventional online bullying and hate speech are primarily perceived not as collective societal harms but rather as isolated individual disputes.⁸¹ Consequently, authorities tend to be more cautious and less aggressive in addressing online hate speech and harassment compared to handling political speech or explicit sexual content.

Both regulators and platforms are largely ineffective in addressing the escalating harms associated with online abuse. Although platforms such as Weibo and Douyin have established community guidelines explicitly categorizing insults, personal attacks, humiliation, and hate speech based on personal characteristics (including birthplace and cultural background) as “harmful information” subject to removal, directly linking specific online comments to offline consequences — such as suicide or self-harm — is exceptionally difficult.⁸² Holding perpetrators accountable is often nearly impossible. Consequently, hate speech typically remains unaddressed until (and even after) it causes tangible, real-world harm.

Even more troublingly, the same regime selectively tolerates or even tacitly encourages certain forms of politically motivated hate speech — particularly nationalist rhetoric aligning with official ideology. For example, the widespread online use of derogatory terms like “Baizuo” (白左, “White Left”), a derogatory term targeting liberal Western values and progressive ideals, remains largely unchecked, implicitly supported by state-driven actors and tolerated by ISPs.⁸³ This selective non-regulation exemplifies the politicized nature of hate speech enforcement in China, where ostensibly neutral rules become instruments of ideological control rather than genuine tools for protecting equality and dignity of people.

2.5.3. AI-Specific Censorship of Hate Speech

Recent AI-specific regulations explicitly integrate those long-standing political and ideological principles into China’s existing framework governing hate speech, likely intensifying the selective enforcement and suppression of speech.⁸⁴ The IMMGAIS explicitly mandate that generative AI must not produce content advocating ethnic hatred, discrimination, or any other prohibited forms of speech. Furthermore, these regulations place stringent obligations on AI developers and providers, requiring them to adopt preventive measures in data selection, algorithm design, model training, and service deployment to prevent discriminatory outcomes based on ethnicity, nationality, religion, gender, age, occupation, or regional origin.⁸⁵

In practical terms, this regulatory environment significantly expands platforms’ and AI providers’ content moderation obligations, embedding politically defined standards of permissible speech within generative AI systems. Thus, providers must proactively implement sophisticated filtering and moderation mechanisms consistent with the state’s ideological preferences to mitigate liability. Because the regulatory priority remains

⁸⁰ IISMM, art 15(4)–(5); PEGNIC, art. 6(6)–(7).

⁸¹ Cao Yin, “Top Court: No Letup in Anti-Cyberbullying Battle,” *China Daily*, March 8, 2024, <https://www.chinadailyhk.com/hk/article/380047>.

⁸² Wang Shuaishuai, “How Hate Speech Falls Through the Cracks of the Chinese Internet,” *Sixth Tone*, November 23, 2022, <https://www.sixthtone.com/news/1011708>.

⁸³ Chen, “Online Hate Speech,” 176–78.

⁸⁴ IMMGAIS, art. 4(1).

⁸⁵ IMMGAIS, art. 4(2); PEGNIC, art. 7.

political conformity rather than genuine anti-discrimination, AI-driven moderation is likely to perpetuate existing biases, exacerbating rather than ameliorating inequalities. Therefore, this ideological conformity potentially causes a chilling effect on legitimate expression, reinforcing self-censorship and limiting the diversity of viewpoints in online discourse.

In sum, China's hate speech regime, augmented by AI-specific regulations, primarily serves state-centric political objectives rather than genuinely addressing discrimination and hate. The selective and politically motivated enforcement of hate speech laws imposes substantial burdens on AI providers, compelling them to implement strict content moderation mechanisms that further entrench political and ideological control over speech in digital spaces.

2.6. Election and Political Content

2.6.1. Disinformation: Legal Framework and Political Underpinnings

In China, the regulation of disinformation (commonly referred to as “rumors” or “false information”) is deeply integrated in the broader structure of speech governance, frequently emphasizing political stability, ideological control, and state-defined public order. Historically, Chinese authorities have employed a combination of criminal law, administrative regulations, and platform self-censorship mechanisms to manage disinformation, especially during sensitive political periods and crises.

Central to this framework is Article 291 of China's Criminal Code, which specifically addresses the creation and dissemination of false information during emergencies. The article imposes criminal penalties on individuals who spread rumors or misinformation, with even harsher sanctions applied when the dissemination results in serious social consequences. Notably, during the COVID-19 pandemic, this provision was widely enforced to manage narratives surrounding public health measures and governmental response, often suppressing legitimate public discourse.⁸⁶ The Wuhan incident in early 2020 — where local authorities detained several bloggers for warning about a SARS-like virus outbreak — epitomizes the tension between state-defined public order and individual freedom of expression.⁸⁷

Moreover, Chinese law enforcement agencies frequently use vaguely defined crimes, particularly “picking quarrels and provoking trouble” (寻衅滋事), as flexible and arbitrary “catch-all” offenses to suppress various forms of expression, including alleged disinformation.⁸⁸ Such vaguely worded charges permit broad prosecutorial discretion, significantly expanding state capacity to control information dissemination without transparent legal standards.⁸⁹

2.6.2. Administrative Regulations and Platform Responsibilities on Disinformation

Beyond criminal sanctions, China employs a robust administrative regime designed explicitly to combat online disinformation. Key regulations include the aforementioned IISMM and PEGNIC. Collectively, these rules require ISPs, online platforms, and algorithmic service providers to proactively identify, filter, and eliminate

⁸⁶ Chen, “Chinese Speech Imperialism,” 523–24.

⁸⁷ Javier C. Hernández, “China Detains Activist Who Accused Xi of Coronavirus Cover-Up,” *New York Times*, February 17, 2020, <https://www.nytimes.com/2020/02/17/world/asia/coronavirus-china-xu-zhiyong.html>.

⁸⁸ Chen, “Chinese Speech Imperialism,” 523.

⁸⁹ Chen, “Chinese Speech Imperialism,” 525.

“rumors” and misinformation that could “disrupt economic and social order,” “disturb social stability,” or “improperly” discuss natural disasters or major accidents.⁹⁰

For example, the PEGNIC explicitly forbids online content producers and users from disseminating information considered “rumors” that disrupt economic and social stability.⁹¹ Platforms must implement stringent measures to prevent and resist the publication of inappropriate commentary on natural disasters or major public incidents, thereby curtailing independent discourse on and analysis of sensitive topics.⁹² Similarly, the IISMM mandates platforms to remove and suppress content classified as rumors or harmful misinformation that threaten public order and stability.⁹³

2.6.3. AI-Specific Regulation of Disinformation

The introduction of AI technologies has led Chinese regulators to expand existing disinformation controls explicitly into generative AI, algorithmic recommendation, and deep synthesis (deepfake) technologies. Under the IMMGAIS, generative AI providers must uphold “socialist core values,” explicitly prohibiting the generation of “false and harmful information.”⁹⁴ This requirement places substantial obligations on developers to ensure that AI-generated content strictly adheres to state-defined accuracy and ideological correctness.

Similarly, the Provisions on the Administration of Algorithm Recommendations for Internet Information Services (PAARIIS) explicitly prohibit providers from using algorithmic recommendations to disrupt economic and social order or to disseminate legally prohibited information, including misinformation.⁹⁵ Providers offering algorithmically generated news content must obtain specific licenses and are explicitly forbidden from creating or synthesizing false news or sharing news from unauthorized sources.⁹⁶ This approach underscores the stringent control exercised by the state over algorithmic news dissemination, which results in significantly restricting independent reporting and public discourse.

The Provisions on the Management of Deep Synthesis of Internet Information Services (PMDSIIS) specifically target deepfake content, prohibiting the creation, duplication, publication, or dissemination of fake news through deep synthesis technology.⁹⁷ Providers are required to implement comprehensive “debunking mechanisms” (辟谣机制) to promptly identify, correct, and report false information.⁹⁸ Additionally, generated or edited deep synthesis content must be prominently labeled to prevent public confusion or misrecognition.⁹⁹

2.6.4. Practical Implications and Selective Enforcement on Disinformation

Despite these extensive regulations, enforcement remains politically selective and strategically instrumental. “Disinformation” that targets government critics or narratives opposing official policy is typically addressed swiftly and harshly.¹⁰⁰ Conversely, misinformation or rumors that support nationalist sentiment or reinforce official positions are frequently tolerated, tacitly endorsed, or even amplified by state-controlled platforms.¹⁰¹

90 IISMM, art. 15(6); PEGNIC, arts. 6(8) and 7(3).

91 PEGNIC, art. 21.

92 PEGNIC, arts. 9–10, 16–17.

93 IISMM, art. 16.

94 IMMGAIS, art. 4(1).

95 PAARIIS, art. 6.

96 PAARIIS, art. 13.

97 PMDSIIS, art. 6.

98 PMDSIIS, art. 11.

99 PMDSIIS, art. 17.

100 Dr. Li Wenliang — Telling People About Covid in Wuhan, Public Security Bureau of Wuhan, Wuchang Division, Zhongnan Precinct, Letter of Reprimand, Wu Public (Central) no. 20200103.

101 Wang Doe — Disrupting Public Order by Posting About Xinjiang, Intermediate People’s Court of Yinchuan, Ningxia Hui Autonomous Region, Administrative Judgment, (2020) Ning 01

Such selective enforcement highlights the inherently political nature of disinformation regulation in China, functioning primarily as a tool of ideological control rather than genuine public interest protection.

Additionally, because criminal and administrative standards lack clear definitions and rely heavily on state-defined interpretations, these rules significantly heighten the risk of arbitrary enforcement. The vague offense of “picking quarrels and provoking trouble,” in particular, likely continues to serve as an omnipresent threat used in AI contexts to suppress critical or inconvenient narratives under the guise of combating disinformation.¹⁰²

In summary, China’s disinformation regulatory framework — further reinforced by explicit AI-specific rules — functions predominantly as a political instrument rather than a neutral mechanism to ensure factual accuracy. AI developers and platforms must navigate complex, often conflicting legal standards that mandate stringent and politically motivated content moderation practices. This regulatory environment places substantial burdens on generative AI and algorithmic content providers, intensifying existing patterns of censorship and self-censorship, thereby restricting the diversity and authenticity of public discourse in China.

2.7. Copyright

Chinese copyright law is structurally different from its counterparts in liberal democracies. Rather than functioning as a purely commercial regime for protecting economic rights, China’s copyright law is embedded in a broader ideological framework. While copyright is traditionally seen in democratic contexts as a catalyst for creativity and free expression, in China it functions as a tool of ideological control and has become a tightly monitored legal domain — repurposed to serve political ends.¹⁰³ As a result, questions surrounding AI-generated content and copyright are not merely technical or economic: they are inherently political, implicating the broader apparatus of censorship and state control over digital expression. This structural difference is key to understanding the role of copyright in China’s AI-related speech regulation.

2.7.1. Censorship as a Precondition for Copyright Protection

Historically, Chinese authorities have used copyright law as a tool to suppress politically sensitive content.¹⁰⁴ Originally, Article 4 of the Copyright Law of the PRC conditioned copyright protection on conformity with state censorship rules. In the 2009 Sino-US copyright dispute, the World Trade Organization found China’s refusal to protect uncensored works inconsistent with its international copyright obligations.¹⁰⁵ Although later revisions to the Copyright Law formally reworded the clause, the requirement that works must not violate the Constitution or harm public interests still functions in practice as a precondition for copyright, effectively excluding politically sensitive or dissenting content from protection. This is most clearly reflected in the institutional arrangement whereby the National Copyright Bureau and the State Administration of Press and Publication operate as a single agency under the framework of “one institution, two nameplates.”¹⁰⁶ In practice, this means that the same officials responsible for copyright enforcement also oversee ideological censorship. Since the 2018 constitutional amendment, the dual-function agency has been formally integrated into the Propaganda Department of the Central Committee of the CCP, reinforcing its role as a central instrument of political control.¹⁰⁷

Administrative Final Instance no. 282, October 14, 2020.

¹⁰² Helen Davidson, “China Should Scrap ‘Picking Quarrels’ Crime, Says Leading Lawyer,” *South China Morning Post*, February 28, 2023, <https://www.theguardian.com/world/2023/feb/28/china-should-scrap-picking-quarrels-says-leading-lawyer>.

¹⁰³ Ge Chen, *Copyright and International Negotiations: An Engine of Free Expression in China?* (Cambridge University Press, 2017), 19–21.

¹⁰⁴ Chen, *Copyright and International Negotiations*, 21–32.

¹⁰⁵ World Trade Organization (WTO) Panel Report, *China — Measures Affecting the Protection and Enforcement of Intellectual Property Rights*, WTO Doc WT/DS362/R, January 26, 2009.

¹⁰⁶ Chen, “Chinese Speech Imperialism,” 533–37.

¹⁰⁷ Chen, “Chinese Speech Imperialism.”

In recent years, China's AI policy and intellectual property (IP) frameworks have converged to reinforce authoritarian governance.¹⁰⁸ This approach is reinforced by policy directives such as the *Outline of National Informatization Development Strategy*, which mandates that internet companies assume primary responsibility for supporting state-led digital governance.¹⁰⁹ For example, during the COVID-19 pandemic, platforms like WeChat actively assisted government surveillance and enforcement measures, illustrating the close integration of corporate operations with state governance objectives.¹¹⁰ In legal practice, existing IP laws have been recalibrated to accommodate AI development.¹¹¹ This convergence enables the state to retain tight ideological control while selectively encouraging technical innovation.

2.7.2. Strict Copyright Liability for Training Data or Outputs

Chinese copyright law includes certain statutory exceptions, but these operate without the constitutional free speech protections found in jurisdictions like the EU or the United States. Overall, China's copyright regime lacks clear, AI-specific liability rules. In practice, developers and users are exposed to significant legal risk even in the absence of intent or negligence, particularly within the discretionary and politically entangled enforcement environment.

There are multiple cases in which users were held liable for copyright infringement under a strict or quasi-objective standard, despite the alleged use being of a non-commercial or incidental nature. These cases serve as compelling examples of how liability may be imposed without requiring proof of fault. Additionally, enforcement in such instances is frequently overshadowed by broader censorship measures, where copyright claims — often in conjunction with moral or reputational concerns — are used to remove or suppress content the state finds objectionable. This convergence of copyright and content regulation creates heightened and unpredictable legal exposure for developers and users of generative AI in China.

For example, the IMMGAIS require that intellectual property rights be respected in the provision and use of generative AI services, including during training processes such as pre-training and optimization.¹¹² Yet these directives remain vague, offering little practical guidance. As a result, Chinese courts have taken the lead, resolving disputes related to AI-generated content using traditional copyright doctrines — especially on questions of authorship and the copyrightability of AI-generated content.¹¹³

Generally, Chinese courts hold that only human creators can claim authorship, since AI models are not recognized as legal entities.¹¹⁴ However, courts have increasingly acknowledged that AI-assisted works can qualify for copyright protection if they embody sufficient human creative input. In *Film v. Baidu* (2018), the Beijing Internet Court held that although the AI-generated analysis did not meet conventional standards for a “work,” it merited some protection due to the collaborative input from both developers and users.¹¹⁵ In *Tencent v. Shanghai Yingxun* (2019), a Shenzhen court upheld the copyright for an article produced by Tencent's AI program “Dreamwriter,” citing the preparatory and editorial decisions made by humans.¹¹⁶

¹⁰⁸ See, e.g., Notice of the State Council on Issuing the Development Plan on the New Generation of Artificial Intelligence [国务院关于印发新一代人工智能发展规划的通知], V. § 3, July 8, 2017. See also Guiding Opinions, arts. II and XIV.

¹⁰⁹ Outline of the National Informatization Development Strategy [国家信息化发展战略纲要], § 52, July 27, 2016.

¹¹⁰ Jing Yang, “WeChat Becomes a Powerful Surveillance Tool Everywhere in China,” *Wall Street Journal*, December 22, 2020, <https://www.wsj.com/articles/wechat-becomes-a-powerful-surveillance-tool-everywhere-in-china-11608633003>.

¹¹¹ Opinions of the Supreme People's Court on Regulating and Strengthening the Applications of Artificial Intelligence in the Judicial Fields [最高人民法院关于规范和加强人工智能司法应用的意见], § 19, August 12, 2022.

¹¹² IMMGAIS, arts. 4(3) and 7(2).

¹¹³ See Zhou Bo, “Artificial Intelligence and Copyright Protection — Judicial Practice in Chinese Courts,” WIPO, accessed July 13, 2025, https://www.wipo.int/about-ip/en/artificial_intelligence/conversation_ip_ai/pdf/ms_china_1_en.pdf.

¹¹⁴ The Copyright Law of the PRC, art. 2, 2020.

¹¹⁵ *Beijing Film Law Firm v. Beijing Baidu Netcom Technology Co., Ltd.*, Beijing Internet Court (2018), Beijing 0491 Min Chu no. 239.

¹¹⁶ *Shenzhen Tencent Computer System Co., Ltd v. Shanghai Yingxun Technology Co. Ltd.*, People's Court of Nanshan (District of Shenzhen) (2019), Yue 0305 Min Chu no. 14010.

More recently, in *Li v. Liu* (2023), the Beijing Internet Court granted copyright protection to an AI-generated image, emphasizing the user’s intellectual contributions — such as crafting prompts and designing stylistic parameters.¹¹⁷ Similarly, in *Wang v. Wuhan X Technology Co. Ltd.* (2024), a Wuhan court ruled that the plaintiff’s use of keywords, light and shadow effects, and creative oversight in an AI-generated image constituted sufficient “personalized expression” to warrant copyright protection.¹¹⁸

2.7.3. Copyright Enforcement Campaigns as a Censorship Tool

Copyright enforcement campaigns — often framed as anti-piracy or anti-infringement drives — serve as an additional censorship mechanism. These campaigns are sometimes deployed to remove or suppress politically sensitive, critical, or parodic works under the guise of copyright protection. Such a move is particularly effective when combined with other legal instruments such as defamation or “rumor-spreading” laws. As such, copyright law enforcement functions as an integral part of China’s broader content control ecosystem, reinforcing ideological conformity and limiting freedom of expression, including in the realm of AI-generated content.

According to the IMMGAIS, all providers of AI-generated content are designated as network information content producers and accordingly must fulfill network information security obligations.¹¹⁹ While this responsibility may seem limited to content integrity or privacy issues, in practice, it extends into copyright enforcement, with explicit censorship implications.

China’s framework for copyright enforcement in the digital environment began with the State Council’s 2006 regulations, which drew inspiration from the US Digital Millennium Copyright Act (DMCA) and introduced a “notice and takedown” regime.¹²⁰ However, unlike in the United States, where the DMCA regime is closely tied to procedural fairness and counter-notice mechanisms, China’s system has evolved into a censorship-enhancing tool. Notably, Chinese law has expanded this regime into a stricter “notice and necessary measures” rule.¹²¹ In the AI context, this mandate means that AI service providers must adopt measures such as deletion, blocking, and link severance when notified of allegedly infringing content — and, crucially, it imposes a heightened duty of care to prevent future violations.¹²²

To be clear, this program of heightened platform liabilities is not confined to copyright law enforcement but, as noted in 2.2.6, is also applicable to the rest of the legal issues discussed in the preceding sections (defamation, explicit content, hate speech, disinformation, etc.). Under this regime, platforms are not only expected to respond reactively to possible infringement violations but are also required to proactively review, filter, and preemptively intercept problematic user-generated content. In recent judicial practice, Chinese courts have increasingly demanded that platforms adopt future-oriented content control mechanisms to prevent potential copyright infringements, where platforms may be held liable if they fail to implement preemptive filtering measures that could have prevented the dissemination of infringing content, even before they receive formal notice.¹²³

¹¹⁷ *Li v. Liu*, Beijing Internet Court (2023), Beijing 0491 Min Chu no. 11279.

¹¹⁸ *Wang v. Wuhan X Technology Co. Ltd.*, People’s Court of Wuhan East Lake New Technology Development Zone (2024), E 0192 Zhi Min Chu no. 968.

¹¹⁹ IMMGAIS, art. 9.

¹²⁰ Regulations on the Protection of the Right of Communication Through Information Networks [信息网络传播权保护条例], arts. 14–15, May 18, 2006.

¹²¹ Civil Code, arts. 1195 and 1197, May 28, 2020.

¹²² IMMGAIS, art. 14.

¹²³ Tang Yili [唐一力], “Research on the Indirect Infringement of the Internet Service Provider — From the Perspective of the Copyright Protection of Major Sports Events as an Example [网络服务提供商间接侵权责任的重新思考—以重大体育赛事节目版权保护为例],” *Legal Forum* [法学论坛] 38, no. 4 (2023): 154–55.

While China's enforcement mechanisms may appear to mirror the sophistication of copyright protection regimes in liberal democracies, they are actually closely integrated with the political censorship infrastructure. Where an alleged infringement is deemed to “damage the public interest,”¹²⁴ a term that remains vague and highly politicized, enforcement is conducted by the courts as well as by copyright administrative authorities that function within the country's propaganda and censorship apparatus.

For example, during the 2021 China Internet Copyright Protection and Development Conference, the deputy minister of the Propaganda Department of the CCP's Central Committee explicitly linked copyright enforcement to “ideological” oversight, calling for intensified copyright law enforcement to rectify “problems strongly complained about by the masses” in areas such as online news, short videos, and livestreams.¹²⁵ Thus, he emphasized the need for stronger responsibility among internet companies to enforce copyright compliance and to bolster “initiative” in managing online content.¹²⁶

Following this directive, the “Sword Net 2021” operation was jointly launched by the National Copyright Administration, Ministry of Industry and Information Technology, Ministry of Public Security, CAC, and other departments.¹²⁷ The campaign targeted public account operators who modified or adapted film and television works into short-form content without authorization and redistributed such works on online platforms, actions that included their unauthorized editing, excerpting, parodying, and uploading of videos.¹²⁸ Many of these activities are often protected by fair use doctrines in the laws of other countries but are treated as copyright infringement in China if they are considered ideological threats.

In sum, copyright law enforcement in the era of AI-generated content continues to serve dual functions: It not only protects proprietary interests but also strengthens state control over digital expression. Under the guise of legal enforcement, Chinese authorities deploy copyright as a means of preemptive censorship — policing not just rights violations but also politically inconvenient speech embedded in user-generated media.

2.8. Measures Empowering Freedom of Expression

Unlike liberal democracies, China does not genuinely pursue policy initiatives explicitly designed to enhance freedom of expression, enable media pluralism, or protect minority voices in its AI governance and broader speech regulatory frameworks. Nevertheless, Chinese authorities occasionally adopt measures or policies that could superficially appear supportive of greater expression, diversity, or digital literacy. Upon closer examination, however, these initiatives invariably serve ideological, political, or national security objectives rather than genuine freedom of speech concerns.

2.8.1. Limited Multilingual and Diversity Initiatives in AI

Occasionally, Chinese authorities highlight the importance of multilingual capabilities and diversity in AI development.¹²⁹ Recent Chinese state-backed initiatives have indeed mandated the expansion of AI

¹²⁴ Copyright Law, art. 53.

¹²⁵ Lai Mingfang [赖名芳], “2021 China Internet Copyright Protection and Development Conference Was Held in Beijing. Zhang Jianchun Attended and Delivered a Keynote Speech [2021中国网络版权保护与发展大会在京召开 张建春出席并作主旨讲话],” *China News Publishing and Broadcasting Newspaper* [中国新闻出版广电报], June 2, 2021.

¹²⁶ Lai, “2021 China Internet Copyright Protection.”

¹²⁷ Wang Jing et al. [王婧等], “Copyright Administration of the Publicity Department of the CPC Central Committee: ‘Sword Net 2022’ Special Action Will Be Launched [中宣部版权管理局: 将启动‘剑网2022’专项行动],” CCTV [央视网], April 26, 2022, https://content-static.cctvnews.cctv.com/snow-book/index.html?toc_style_id=feeds_default&share_to=wechat&item_id=1043455964018346083&track_id=35EE615C-7772-43FD-B704-1E5A1328F236_672652686231.

¹²⁸ Wang et al., “Copyright Administration.”

¹²⁹ Notice of the State Council on Issuing the New Generation Artificial Intelligence Development Plan [国务院关于印发新一代人工智能发展规划的通知], § III.(2), July 8, 2017.

capabilities into multiple languages, including less commonly used or minority languages within the PRC.¹³⁰ On the surface, such efforts may even appear consistent with global best practices encouraging AI-driven language diversity and, thus, expressive diversity. However, the underlying objective here is predominantly ideological rather than supportive of authentic diversity.¹³¹ Enhancing multilingual capacities in AI facilitates the global dissemination of CCP-approved narratives, thereby reinforcing state ideological governance among diverse linguistic and ethnic groups both domestically and internationally.¹³²

2.8.2. AI Literacy and Public Education from a National Security Perspective

Chinese authorities regularly engage in public education and literacy campaigns surrounding digital and AI technologies.¹³³ However, such initiatives emphasize “national security” and “social stability,”¹³⁴ framing digital literacy explicitly as a means to protect citizens from perceived ideological and external threats rather than to foster informed critical engagement. Official education materials and initiatives often reinforce state-defined boundaries of acceptable discourse, instructing citizens on compliance with ideological norms, identification of “rumors,” and avoidance of “misinformation.”¹³⁵ Thus, rather than empowering citizens to critically navigate information freely, Chinese AI literacy programs function primarily to reinforce ideological conformity and state-defined information control.

2.9. Miscellaneous

Significantly, China’s AI strategy explicitly integrates a transnational dimension into its broader regulatory framework, embracing an “AI sovereignty” model that seeks to export authoritarian speech regulation standards globally.¹³⁶ Central to this approach is China’s Global Artificial Intelligence Governance Initiative, proposed by Xi Jinping at the Belt and Road Initiative Forum in 2023. The initiative promotes a vision of AI governance prioritizing “national sovereignty,”¹³⁷ thereby legitimizing stringent control and state oversight over AI-generated content.

2.9.1. From Defensive Censorship to Offensive Regulation

Traditionally, China’s strategy of speech control emphasized defensive mechanisms — namely, censorship — to control and eliminate speech deemed harmful to state interests. Recently, however, the CAC has explicitly adopted an offensive strategy in content governance: promoting state-approved narratives domestically and exporting positive portrayals of China internationally.¹³⁸ This shift was notably institutionalized in the 2019 PEGNIC, which — for the first time — legally encouraged network information producers to create and disseminate content that explicitly serves CCP ideological goals.¹³⁹ Such content includes promoting Xi Jinping Thought, actively disseminating the CCP’s political doctrines and policies, highlighting China’s

130 Opinions of the Ministry of Education, the National Language Commission, and the Central Cyberspace Affairs Commission on Strengthening the Construction of Digital Chinese and Promoting the Development of Language and Writing Informationization [教育部 国家语委 中央网信办关于加强 数字中文建设 推进语言文字信息化发展的意见], § IV (8), January 13, 2025 (hereafter cited as 2025 Opinions).

131 2025 Opinions, § I.

132 Richard Heeks et al., “China’s Digital Expansion in the Global South: Systematic Literature Review and Future Research Agenda,” *The Information Society: An International Journal* 40, no. 2 (2024): 69, <https://doi.org/10.1080/01972243.2024.2315875>.

133 Liu Caiyu, “Chinese Education Ministry Proposes AI Integration into School Curricula, Teaching Materials,” *Global Times*, April 16, 2025, <https://www.globaltimes.cn/page/202504/1332246.shtml>.

134 2025 Opinions, § III (7).

135 “China Launches Online Platform to Combat Education Sector Rumors,” *Xinhua*, March 21, 2024, https://english.www.gov.cn/news/202403/21/content_WS65fbc738c6d0868f4e8e54f3.html.

136 Matthew J. Dagher-Margosian, “CCP Cyber Sovereignty Contains Lessons for AI’s Future,” *James Town Foundation China Brief*, April 12, 2024, <https://jamestown.org/program/ccp-cyber-sovereignty-contains-lessons-for-ais-future/>.

137 Cao Desheng, “China Committed to Actively Promoting Development, Governance of AI,” *China Daily*, February 13, 2025, <https://asianews.network/china-committed-to-actively-promoting-development-governance-of-ai/>.

138 Chen, “Chinese Speech Imperialism,” 488–90.

139 PEGNIC, art. 5.

social and economic achievements, advancing socialist core values, and reinforcing consensus around CCP positions.¹⁴⁰ Specifically, Article 5 of the PEGNIC encourages online content that showcases China’s cultural prestige internationally by generating influential narratives and presenting a “true, three-dimensional, and comprehensive” image of China globally.¹⁴¹

Moreover, the PEGNIC requires online platforms to prioritize these government-endorsed narratives across virtually all prominent digital spaces, including news home pages, search engines, recommendation algorithms, social media trending topics, and entertainment and e-commerce platforms.¹⁴² As a result, the Chinese online ecosystem has increasingly become a managed environment designed expressly to cultivate state-approved speech, crowding out alternative viewpoints and implicitly constraining free and diverse public expression.¹⁴³

2.9.2. Case Study: DeepSeek and Algorithmic Propaganda

The implications of China’s offensive speech regulation are vividly illustrated by the generative AI chatbot DeepSeek, developed by a Chinese AI company. DeepSeek has recently gained international attention as an AI system that not only employs stringent censorship of politically sensitive topics but actively generates responses echoing official Beijing talking points, serving as a significant tool of algorithmically driven propaganda.

A high-profile analysis revealed that DeepSeek consistently produced responses reflecting official Chinese geopolitical positions, echoing narratives favorable to Beijing on contentious international issues such as Taiwan, Hong Kong, Xinjiang, and global governance.¹⁴⁴ Such algorithmically generated content functions not merely as passive censorship but as proactive ideological influence, facilitating state-driven narratives’ ability to reach international audiences in sophisticated ways.¹⁴⁵

Further investigations highlighted the specific mechanics of DeepSeek’s censorship practices, demonstrating the AI’s nuanced capacity to selectively suppress or reshape responses according to politically sensitive criteria.¹⁴⁶ Although some users have attempted workarounds to evade DeepSeek’s ideological filters, the inherent structure and training methodologies of the AI ensure adherence to state-defined mainstream values, illustrating the opportunities for deeper integration of censorship and propaganda into generative AI technologies.¹⁴⁷

2.9.3. International Backlash and Regulatory Scrutiny

The offensive nature of DeepSeek’s state-oriented outputs has prompted international concern, leading multiple countries to object to its use on grounds of national security, misinformation, undue political influence, and privacy/data protection. The United States, South Korea, Italy, Australia, and Taiwan have already banned or significantly restricted access to DeepSeek, citing risks posed by its embedded ideological bias and censorship.¹⁴⁸ Meanwhile, France, Belgium, and Ireland have launched formal investigations into the

¹⁴⁰ PEGNIC, art. 5(1)–(5).

¹⁴¹ PEGNIC, art. 5(6).

¹⁴² PEGNIC, art. 11.

¹⁴³ Valentin Weber, *Data-Centric Authoritarianism: How China’s Development of Frontier Technologies Could Globalize Repression* (National Endowment for Democracy, February 11, 2025), 8, https://www.ned.org/wp-content/uploads/2025/02/NED_FORUM-China-Emerging-Technologies-Report.pdf.

¹⁴⁴ Jordi Calvet-Bademunt et al., “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>.

¹⁴⁵ “We Asked DeepSeek About Geopolitics: It Gave Us Beijing Talking Points,” *Politico*, February 4, 2025, <https://www.politico.eu/article/we-asked-deepseek-about-geopolitics-chinese-government-propaganda-artificial-intelligence/>.

¹⁴⁶ Zeyi Yang, “Here’s How DeepSeek Censorship Actually Works — and How to Get Around It,” *The Wired*, January 31, 2025, <https://www.wired.com/story/deepseek-censorship/>.

¹⁴⁷ US Select Committee on the CCP, *DeepSeek Unmasked*, 4–6.

¹⁴⁸ Pascale Davies, “DeepSeek: Which Countries Have Restricted the Chinese AI Company or Are Questioning It?,” *Euronews*, February 3, 2025, <https://www.euronews.com/next/2025/02/03/>

implications of using DeepSeek and similar AI technologies originating in authoritarian contexts, underscoring the growing international unease about China's AI-driven propaganda capabilities.¹⁴⁹ Additionally, DeepSeek faces imminent bans in Germany from leading manufacturers such as Apple and Google due to concerns about data protection.¹⁵⁰

This international response highlights the global implications of China's strategic shift toward offensive speech regulation via algorithmically powered propaganda. Whereas traditional censorship was primarily inward-facing and designed to silence domestic dissent, contemporary regulatory frameworks harness AI's capabilities to project ideological influence outward, actively reshaping international perceptions and public discourses around sensitive political topics.

deepseek-which-countries-have-restricted-the-chinese-ai-company-or-are-questioning-it.

¹⁴⁹ Davies, "DeepSeek."

¹⁵⁰ Hakan Ersen and Miranda Murray, "DeepSeek Faces Ban from Apple, Google App Stores in Germany," *Reuters*, June 27, 2025, <https://www.reuters.com/sustainability/boards-policy-regulation/deepseek-faces-expulsion-app-stores-germany-2025-06-27/>.

3. Conclusion

This chapter has analyzed the extensive impact of China's AI policy and legal framework on freedom of expression, illuminating the systematic incorporation of AI technologies into China's broader regime of speech regulation. Rather than genuinely safeguarding expressive freedoms, China's governance model emphasizes state control, ideological conformity, and stringent censorship.

AI-specific legislation and policies amplify traditional restrictions on speech in the PRC. China's regulatory landscape, marked by core laws such as the Cybersecurity Law, Data Security Law, and Personal Information Protection Law, explicitly embeds strict national security and ideological compliance into data governance and AI development. The resulting regulatory environment has introduced significant obligations for AI providers, mandating proactive censorship of politically sensitive content across categories, including copyright, defamation, explicit content, hate speech, and disinformation. Consequently, AI-generated content in China encounters unprecedented scrutiny, with developers and users facing severe legal and political risks.

The integration of copyright law into AI governance exemplifies how legal regimes nominally designed to encourage innovation have been repurposed as tools for political censorship. Although recent judicial decisions appear to provide limited recognition for human inputs into AI-generated works, these decisions remain fundamentally constrained by censorship imperatives embedded in China's dual-purpose copyright administration. Similarly, defamation and libel laws employ expansive standards that criminalize online speech based on quantifiable dissemination metrics, significantly restricting critical discourse.

In addressing explicit content, Chinese regulations broadly and vaguely define obscenity and political pornography, strategically conflating sexual and political expression to facilitate extensive censorship. Rules targeting AI-driven hate speech and disinformation similarly prioritize regime stability and ideological uniformity over authentic protection of equality or factual accuracy, thereby intensifying state oversight of online speech through AI-enabled surveillance and censorship mechanisms.

Notably absent from China's AI governance framework are genuine initiatives to empower free expression. Even limited measures ostensibly promoting multilingualism or AI literacy are framed within strict ideological and national security narratives.

Rather than ensuring freedom of expression, China's approach instrumentalizes AI and digital platforms for state propaganda. By mandating and incentivizing content creation that explicitly supports the government ideology, China's regulatory environment constrains authentic diversity and plurality of expression. The case of DeepSeek clearly demonstrates the tangible international ramifications of such offensive speech regulation, revealing a strategic shift toward global ideological influence and underscoring the urgent need for international scrutiny of algorithmically driven propaganda originating in authoritarian contexts.

Globally, China’s assertive propagation of its “AI sovereignty” model through initiatives like the Global Artificial Intelligence Governance Initiative presents profound challenges to international norms of free expression and democratic governance. China’s strategic integration of censorship and propaganda into generative AI technologies signals an alarming precedent, particularly as its influence expands via platforms such as DeepSeek. This approach starkly contrasts with democratic frameworks around emerging technologies, which prioritize transparency, accountability, and pluralism in AI governance.

Looking forward, China’s evolving AI regulatory landscape will likely continue to solidify state control and ideological conformity, exerting increasing pressure internationally as Chinese-developed AI systems and governance standards spread globally. This trend calls for heightened vigilance, proactive policy responses, and coordinated international efforts to ensure that global AI governance remains anchored in commitments to human rights, democratic values, and genuine freedom of expression.



**THE
FUTURE
OF
FREE
SPEECH**

OCTOBER 2025

www.futurefreespeech.org