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.

© The Future of Free Speech 2025
Publication Date: October 2025

Preface

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

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

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

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

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

- AI legislation in Brazil, China, the European Union, India, the Republic of Korea, and the United States.[9] In this report, AI legislation refers to laws and public policies addressing AI-generated content, with particular focus on elections and political speech, hate speech, defamation, explicit content (including

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

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. 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 Brazil

Carlos Affonso Souza*

*Carlos Affonso Souza is a professor at the State University of Rio de Janeiro (UERJ). He holds a PhD (2009) and a master's degree (2003) in Private Law from UERJ and is a director of the Institute for Technology and Society (ITS Rio), a leading organization in Brazil focusing on tech policy and regulation. Souza was one of the main contributors in the creation of Brazil's Internet Bill of Rights (2014) and is currently involved in the debates concerning data protection and AI regulation. He is a visiting professor at the University of Ottawa Law School and an affiliated fellow at the Information Society Project/Yale University Law School. He writes weekly about law and technology for UOL, the largest Brazilian online news outlet.

Abstract

In this chapter we analyze how generative artificial intelligence (AI) is being regulated in Brazil, focusing on its impact on freedom of expression. We explore the country's constitutional protections for expression, the emerging legislative framework — including the Artificial Intelligence Bill (PL 2338/2023) — and how sector-specific policies intersect with AI regulation. The chapter examines issues such as liability for AI-generated content and restrictions related to copyright, defamation, hate speech, and disinformation, as well as how the regulation of high-risk AI systems, if not properly balanced, could affect journalistic, artistic, and political speech. The bill introduces categorical prohibitions on certain uses of AI and imposes governance requirements on generative AI. We conclude by identifying opportunities and challenges in ensuring that AI development in Brazil remains aligned with democratic values and provides robust protections for freedom of expression.

Carlos Affonso Souza

Carlos Affonso Souza is a professor at the State University of Rio de Janeiro (UERJ). He holds a PhD (2009) and a Master's (2003) degree in Private Law (UERJ) and is a Director of the Institute for Technology and Society (ITS Rio), a leading organization in Brazil focusing on tech policy and regulation. Professor Souza was one of the main contributors in the creation of Brazil's Internet Bill of Rights (2014) and is currently involved in the debates concerning data protection and AI regulation. He is a Visiting Professor at the University of Ottawa Law School and an Affiliated Fellow at the Information Society Project/Yale University Law School. He writes weekly about law and technology for UOL, the largest Brazilian online news outlet.

1. Introduction

Brazil occupies a unique position in global debates on digital rights, often balancing progressive legal frameworks with a complex political environment. The Brazilian Constitution guarantees freedom of expression in broad terms and has become a reference point for internet regulation, particularly throughout the country's Internet Bill of Rights (Marco Civil da Internet, or MCI), a federal law approved in 2014 after an online public consultation.[1] However, the increasing use of generative AI poses novel regulatory and normative challenges.

Brazil's legislative efforts have culminated in the recent approval of a bill (PL 2338/2023) by the Senate, which aims to create a national AI governance framework. The bill adopts a risk-based regulatory model, introduces obligations for transparency, and defines responsibilities across the AI value chain. It also recognizes freedom of expression as a core principle of the law — an acknowledgment of the tension between regulating AI harms and preserving democratic communication.

Here we investigate how freedom of expression interacts with Brazil's existing legal framework and the proposed AI regulations. We do so through the lens of constitutional law, international human rights obligations, and thematic areas such as copyright, defamation, and disinformation. The aim is to clarify how Brazil is shaping its AI governance model and to assess whether it strengthens or threatens expressive freedoms in the digital age.

1 For more information on the public consultation process and the contributions of different stakeholders: Carlos Affonso Souza, Fabro Steibel, and Ronaldo Lemos, "Notes on the Creation and Impacts of Brazil's Internet Bill of Rights," *Theory and Practice of Legislation* 5 (2017): 73-94, https://doi.org/10.1080/20508840.2016.1264677. See also Daniel Arnaudo, "Brazil, the Internet and the Digital Bill of Rights: Reviewing the State of Brazilian Internet Governance," *Instituto Igarapé*, accessed September 14, 2025, https://igarape.org.br/marcocivil/en.

2. Substantive Analyses

2.1. General Standards of Freedom of Expression

Brazil's constitutional and legal framework offers strong protections for freedom of expression. The 1988 Federal Constitution states that "the expression of thought is free, and anonymity is forbidden."[2]

This provision sits within a broader set of fundamental rights that include access to information, freedom of the press, and artistic, scientific, and communicative freedom.[3] These protections are reinforced by Brazil's international commitments, particularly under the American Convention on Human Rights (ACHR), to which Brazil is a party and which recognizes freedom of thought and expression as a cornerstone of democratic society.[4]

Historically, Brazil's Supreme Court (Supremo Tribunal Federal, or STF) has championed a robust interpretation of expressive freedom. In the seminal ADPF 130 case, the court struck down the Press Law enacted during Brazil's military dictatorship, ruling that freedom of expression occupies a "preferential position" within the constitutional order. In doing so, it emphasized that censorship, prior restraints, and disproportionate liability frameworks are incompatible with democratic values.[5]

Yet this strong jurisprudence has faced new pressures in the digital era. Particularly since the January 8, 2023, attacks on Brazil's democratic institutions, including the Supreme Court itself, the STF has adopted more nuanced positions in online speech cases. Under the leadership of Justice Alexandre de Moraes, the court has ordered the removal of social media accounts and, in more extreme cases, the blocking of entire platforms, such as the temporary suspension of X (formerly Twitter). These measures have sparked national and international debate, raising questions about proportionality, due process, and the compatibility of such actions with international standards on freedom of expression. While the court has justified these decisions as necessary to protect democratic order and prevent the spread of harmful content, critics argue that they mark a departure from the STF's traditional speech-protective stance.[6]

2 This wording, from Article 5, item IX, of the Federal Constitution, is often cited in debates over online speech, especially in relation to anonymous or pseudonymous accounts on digital platforms. While the text may suggest a blanket prohibition of anonymous expression, the Brazilian Supreme Court (STF) has interpreted it more narrowly. In a leading opinion by Justice Celso de Mello, the court clarified that the constitutional ban on anonymity does not require prior identification for speech to be lawful; rather, it ensures that mechanisms exist to identify the speaker post hoc in case of violations of third-party rights, such as defamation or incitement. The principle is one not of mandatory real-name attribution but of accountability. See STF, *Mandado de Segurança* No. 24.369 MC/DF, Justice Celso de Mello, October 10, 2002, (Braz.).

3 *Constituição da República Federativa do Brasil de 1988* (Braz. Const.), art. 5, IV, IX, XIV (1988).

4 Organization of *American States, American Convention on Human Rights "Pact of San José, Costa Rica,"* November 22, 1969, art. 13.

5 STF, *Arguição de Descumprimento de Preceito Fundamental* (ADPF) No. 130, Justice Ayres Britto, April 30, 2009 (Braz.).

6 Jack Nicas and André Spigariol, "To Defend Democracy, Is Brazil's Top Court Going Too Far?," *New York Times,* September 26, 2022, https://www.nytimes.com/2022/09/26/world/americas/bolsonaro-brazil-supreme-court.html.

The jurisprudential tension grew in importance when, in June 2025, the Supreme Court decided that Article 19 of Brazil's MCI, which provided a safe harbor for internet platforms from liability for third-party content unless there is a judicial takedown order, was partially unconstitutional. The case set the stage for the STF to adopt a more interventionist posture in light of growing concerns about online harms.[7]

This ruling, along with several others requiring US-based social media companies to remove content or block accounts, including those of Brazilian users operating in the United States, has raised concerns among US authorities. In response, an executive order was issued in connection with the increase of tariffs on Brazilian goods and services exported to the United States.[8]

These shifting judicial waters intersect with the emergence of generative AI, which challenges traditional frameworks for authorship, liability, and intent. While current jurisprudence does not yet directly address AI-generated content, Brazil's broader legal framework provides a normative baseline: Expression should be protected unless it directly infringes upon other rights or legal interests. The difficulty lies in drawing that line when the "speaker" is no longer human. In addition, AI can widely spread fabricated content that, at a first glance, seems authentic.

Notably, the AI Bill (PL 2338/2023) incorporates freedom of expression as a central principle[9] and includes the concept of "integrity of information" as a means to strengthen rather than curtail expressive rights.[10] The law also introduces new obligations related to synthetic content and expands due process guarantees for individuals affected by automated decisions.

In sum, Brazil's legal tradition strongly supports freedom of expression, but recent jurisprudential developments, especially in the context of digital platforms, suggest a more fluid and contested landscape. The regulation of AI-generated content will unfold within this evolving framework, and much will depend on how courts reconcile the promise of technological innovation with the imperatives of democratic accountability and rights protection.

2.2. AI-Specific Legislation and Policies

Brazil is in the process of defining a comprehensive national framework for artificial intelligence through the AI Bill (PL 2338/2023), which has already been approved by the Federal Senate. The bill represents Brazil's most ambitious attempt to regulate AI and includes specific provisions aimed at generative and general-purpose systems. Heavily inspired by the European-style precautionary model, the Brazilian approach also contains some peculiarities, such as the introduction of a chapter focusing on rights granted to those who are "affected by AI."

7 Pedro de Perdigão Lana, Flavio Rech Wagner, and Paulo Rena da Silva Santarém, "Internet Impact Brief — Proposals to Regulate Content Moderation on Social Media Platforms in Brazil," *Internet Society*, March 13, 2022, https://www.internetsociety.org/wp-content/uploads/2022/07/External-IIB-Content-Moderation-Brazil.pdf.

8 According to the executive order: "Indeed, certain Brazilian officials have issued orders to compel United States online platforms to censor the accounts or content of United States persons, where such accounts or content are protected by the First Amendment to the United States Constitution within the United States; block the ability of United States persons to raise money on their platforms; change their content moderation policies, enforcement practices, or algorithms in ways that may result in the censorship of the content and accounts of United States persons; and provide the user data of accounts belonging to United States persons, facilitating the targeting of political critics in the United States." Exec. Order No. 14323, 90 FR 37739 (July 30, 2025), https://www.federalregister.gov/d/2025-14896.

9 Bill No. 2338/2023, "Development, Fostering, and Responsible Use of Artificial Intelligence," art. 2, III (December 10, 2024) (Braz.).

10 Bill No. 2338/2023, art. 2, XV (2024).

2.2.1. Risk-Based Approach

PL 2338/2023 is structured around a risk-based regulatory framework. It categorizes AI systems into three broad levels: prohibited (excessive risk), high risk, and low or undefined risk. The AI Bill also provides a definition for "systemic risk" as "potential negative effects arising out of general-purpose or generative AI systems with relevant impacts on individual and social fundamental rights".[11] Among those deemed "excessive risk" and therefore banned are systems that exploit human vulnerabilities, score citizens based on social behavior, or enable mass biometric surveillance in public spaces without strict judicial oversight.[12] Generative AI systems, depending on their function and impact, may fall under either the high-risk or systemic-risk categories, particularly when deployed in areas such as education, health, and employment.[13] These systems require algorithmic impact assessments, human oversight, and robust documentation throughout the AI life cycle.

2.2.2. General-Purpose and Generative AI

The AI Bill introduces tailored obligations for developers of general-purpose and generative AI systems. It defines "general-purpose AI" as systems trained on large datasets capable of performing a wide range of tasks, and "generative AI" as those that significantly create or alter text, images, audio, video, or code.[14] When such systems are deemed to pose "systemic risks" to fundamental rights, the environment, or democratic processes, they are subject to enhanced transparency and safety obligations.[15]

For generative AI specifically, developers must conduct preliminary risk assessments; ensure datasets are lawfully acquired; disclose summaries of training data; implement environmental sustainability standards; document model behavior and instructions for deployment; and label synthetic content with appropriate identifiers, especially when the output could be confused with authentic human expression.

Artistic, cultural, and entertainment uses are explicitly protected: When content is clearly fictional and does not risk deceiving the public, disclosure requirements may be satisfied through nonintrusive means, such as metadata or credits.[16]

2.2.3. Open-Source vs. Proprietary Models

Brazil's AI Bill does not treat open-source models as exempt from regulation, but it does recognize the need for differentiated treatment. PL 2338/2023 allows for regulatory simplification for systems developed in open and noncommercial environments, especially during the research and development phase.[17] However, once placed into the market or used in real-world conditions, even open-source models may trigger risk-based obligations. For example, a large language model (LLM) released under an open license but deployed in high-risk domains — such as health care or electoral systems — must comply with documentation, impact assessments, and transparency requirements.

11 Bill No. 2338/2023, art. 3, XXX (2024).
12 Bill No. 2338/2023, art. 13 (2024).
13 Bill No. 2338/2023, art. 14; art. 15, VII (2024).
14 Bill No. 2338/2023, art. 4, III–IV (2024).
15 Bill No. 2338/2023, arts. 29–33 (2024).
16 Bill No. 2338/2023, art. 19, §3º (2024).
17 Bill No. 2338/2023, art. 1, §1º; art. 73 (2024).

A notable state-level development is the approval of the Law for Promoting Innovation in Artificial Intelligence in the state of Goiás, the first comprehensive AI statute enacted in Brazil. Goiás adopts a pro-open-source posture, mandating preferential use of open-source software and models in all public-sector AI deployments unless a technical justification is provided. It also institutes an open AI innovation program with financial incentives, public-private partnerships, and awards to recognize impactful use of open and auditable models.[18]

The AI Law of Goiás emphasizes code transparency and auditability. It frames open-source development not only as a tool for innovation but also as a guarantee for sovereignty, competitiveness, and public oversight. Whereas PL 2338/2023 provides regulatory relief for open-source projects during R&D, the Goiás law creates institutional preferences and structural incentives for open models at all stages of development and deployment, with detailed rules for regulatory sandboxes. It even creates a state computing infrastructure to support training and access to high-performance computing for smaller developers using open-source models.

This divergence between federal and state-level initiatives highlights the potential for multilevel AI governance in Brazil, with subnational units like Goiás acting as experimental laboratories. If upheld legally and supported by institutional mechanisms, such state laws could push the national debate forward — especially in the direction of transparency, accessibility, and local innovation ecosystems.

2.2.4. Accountability Across the AI Value Chain

PL 2338/2023 introduces a detailed allocation of responsibilities across the AI value chain — developers, distributors, and deployers *(aplicadores)* — each of whom may be held accountable based on their role and the knowledge they have about the system's use.[19] This distributed model of responsibility is meant to prevent the dilution of liability that often occurs in complex digital ecosystems. When harm arises, courts are expected to evaluate the agent's diligence, risk mitigation efforts, and degree of control over the AI's operation.

The bill includes a safeguard allowing courts to reverse the burden of proof in civil liability cases when the technical opacity of an AI system would make it unreasonably difficult for a harmed individual to meet their evidentiary burden.

The law also explicitly preserves the application of Brazil's Consumer Protection Code and Civil Code, reinforcing that AI is not beyond the reach of existing liability frameworks.

18 Legislative House of the State of Goiás, "Establishes the State Policy for the Promotion of Innovation in Artificial Intelligence in the State of Goiás, https://legisla.casacivil.go.gov.br/api/v2/pesquisa/legislacoes/110694/pdf.
19 Bill No. 2338/2023, art. 4, V-VIII; art. 18 (2024).

2.3. Defamation

Brazilian law provides protections for honor and reputation through both criminal and civil liability mechanisms. The Penal Code criminalizes *calúnia* (false accusation of a crime), *difamação* (false statements that damage reputation), and *injúria* (insults to dignity or decorum). In parallel, the Civil Code and the Consumer Protection Code (CDC) provide for civil liability, including compensation for moral damages arising from content.

These frameworks were developed in a human-centric legal context but are now being tested by the emergence of AI-generated speech, where defamatory outputs may originate from LLMs without direct human authorship or intent.

2.3.1. Liability for AI-Generated Defamation

Under traditional doctrine, intent or fault is a precondition for criminal defamation. Since AI systems lack mens rea, criminal sanctions are unlikely to apply directly to outputs from LLMs. In civil law, however, the landscape is more complex. Brazilian law allows for both fault-based and strict liability regimes depending on context.

The Civil Code establishes that anyone who engages in a risky activity and causes harm must compensate for damages regardless of fault. This provision introduces strict liability in cases involving heightened risk.[20] The CDC similarly holds suppliers strictly liable for damages caused by defects in products or services, even when there is no intent or negligence.

As such, if a generative AI tool is marketed to consumers and produces defamatory content, strict liability could be invoked on the basis that the harm arises from a defective or risky service. Courts may consider whether the output was foreseeable, preventable, or linked to insufficient safeguards in the AI's design or deployment.

This is particularly relevant given that PL 2338/2023 adopts a risk-based classification of AI systems. While the bill reaffirms that the existing liability regimes in the Civil Code and CDC remain in force, it also introduces important procedural innovations:

- Courts may reverse the burden of proof in civil cases where AI opacity prevents the injured party from establishing causation.
- Judges may use the risk categorization of an AI system — as defined under PL 2338/2023 — to determine whether strict liability should apply, even if the defendant argues for a fault-based regime.

In practice, this opens the door for courts to recalibrate liability depending on the AI's classification under PL 2338/2023: High-risk or systemic-risk applications are more likely to trigger strict liability, whereas general-use or low-risk systems may benefit from traditional negligence standards.

20 Law No. 10.406, Civil Code, art. 927, sole paragraph (January 10, 2002) (Braz.).

2.3.2. Intermediary Liability and the Role of Article 19 of the MCI

Brazil's Internet Bill of Rights, or MCI (officially Law No. 12.965/2014), establishes a regime of safe harbor for intermediaries, shielding platforms from liability for third-party content unless they fail to comply with a specific judicial order for removal (Article 19). This model, partially inspired by US Section 230 and European notice-and-takedown mechanisms, has underpinned Brazil's platform regulation for over a decade. However, as previously mentioned, a recent decision by the Supreme Court rendered this provision partially unconstitutional but maintained its enforcement for defamation cases.

A crucial distinction must be made here: LLM-generated content is not third-party content in the traditional sense. When a social media platform like Instagram or X hosts a user's post, it is facilitating publication. In contrast, when a company's AI model (e.g., a chatbot or generative assistant) produces text, the content is generated natively by the system — often based only on minimal prompting.

In such cases, plaintiffs may argue that the output represents the company's own speech or product function, not a third-party contribution. This removes the protection of Article 19 and may expose providers to direct liability for defamatory AI outputs.

2.3.3. Emerging Framework Under PL 2338/2023

PL 2338/2023 attempts to strike a balance between innovation and accountability. It does not displace the current fault/strict liability distinction but creates the conditions for courts to operationalize risk grading as a gateway to liability regime selection, as previously mentioned. This structure allows judges to ask questions such as: Is the system classified as high or systemic risk? Could the harm have been reasonably anticipated? Did the developer or deployer implement appropriate safeguards?

If the answers to these questions point toward elevated risk, courts may impose strict liability even absent fault, in line with both the Civil Code and consumer protection jurisprudence.

This is a subtle but significant shift. While PL 2338/2023 stops short of imposing strict liability across the board, it effectively codifies a risk-informed path to strict liability. As the judiciary confronts more AI-generated speech cases, we are likely to see risk classifications, system transparency, and deployment context play a central role in shaping outcomes — particularly in high-stakes scenarios involving reputational harm and personality rights.

2.4. Explicit Content

Brazil's legal framework includes specific provisions to address the creation, dissemination, and removal of sexually explicit content, particularly when such content is produced or shared without consent. The use of generative AI to fabricate or manipulate intimate imagery has introduced new layers of complexity, as it challenges traditional definitions of authorship, intent, and consent in the digital environment.

2.4.1. Legal Protections Against Nonconsensual Intimate Content

A cornerstone of Brazil's legal response to this issue is Law No. 13.772/2018, which amended the Penal Code to criminalize the unauthorized production or dissemination of nude or sexual images. The Penal Code establishes penalties for both the original recording and for montages or fabrications — a category that directly encompasses deepfake pornography.[21] This provision makes it clear that nonconsensual image generation, including by artificial means, is punishable regardless of whether the depicted scene ever occurred in reality.

This means that if a generative AI model is used to insert someone's likeness into explicit content, it could fall under the scope of Article 216-B, especially the sole paragraph, which criminalizes fabrications where "a person is inserted into a scene of nudity or sexual act."

The use of generative AI to create synthetic nonconsensual intimate images may also give rise to civil liability for moral damages, especially under Brazil's standards for dignity violations and emotional harm, both usually broadly framed and inconsistently applied by the judiciary.

2.4.2. Article 21 of the MCI: A Notice-and-Takedown Mechanism

Beyond criminal and civil sanctions, Brazil's Internet Bill of Rights offers a specialized mechanism to address this type of content. Article 21 of the MCI introduces a notice-and-takedown regime for explicit material involving nudity or sexual acts disseminated without the consent of the participant(s).

Under Article 21, individuals affected can request removal directly from the platform, without prior judicial authorization. Once notified of the problematic content, the provider must act "diligently" to make the content unavailable, or else they may be held liable for the resulting harm. The rule is limited to images, videos, or other materials that depict nudity or sexual acts and must involve identifiable persons.

This provision has been broadly applied in practice and is particularly relevant for AI-generated deepfakes that place real individuals into synthetic adult scenes. Even if the image is fictional, courts have generally upheld Article 21's applicability where the person is clearly recognizable and did not consent to the representation.

Importantly, this provision is not limited to content created by humans. As AI-generated explicit content becomes more prevalent, this mechanism is likely to be invoked more frequently, and platforms will be expected to respond swiftly to takedown requests, regardless of the content's synthetic origin.

2.4.3. Generative AI and PL 2338/2023

PL 2338/2023 addresses these issues indirectly but meaningfully. It prohibits AI systems that facilitate the production or dissemination of child sexual abuse material (CSAM),[22] classifying such systems as involving excessive risk. Moreover, all generative systems are required to include identifiers in synthetic content to verify its provenance. This identification obligation is key in distinguishing fabricated from authentic media, particularly in contexts involving reputational or sexual harm.

21 Decree-Law No. 2.848, Penal Code, art. 216-B (December 7, 1940) (Braz.).
22 Bill No. 2338/2023, art. 13, I, d (2024).

The AI Bill further mandates collaboration between public and private actors to promote the capacity to detect and trace synthetic content. This could facilitate early identification of AI-generated explicit media and support rapid takedown across platforms.

Developers and deployers who fail to implement preventive measures or who ignore signals of abuse could be held accountable under the general liability principles of PL 2338/2023. When used in high-risk contexts, these systems must undergo algorithmic impact assessments,[23] including consideration of how they may be misused to produce sexually explicit or intimate content.

2.4.4. Enforcement and Future Trends

While Brazil's current criminal and civil laws offer a robust framework to address nonconsensual intimate imagery, enforcement still depends heavily on user complaints and platform responsiveness. The presence of Article 21 as a direct takedown route is a critical tool, but its scope is limited to content involving nudity or sexual acts. Other types of synthetic harm (e.g., AI-generated harassment or impersonation without nudity) may not benefit from the same expedited protection.

Additionally, there remains legal uncertainty about who is liable for AI-generated explicit content: Is it the developer, the deployer, or the user? PL 2338/2023's multi-agent liability framework allows courts to assign responsibility across the AI life cycle, depending on who had control or foreseeability of the harm. In practice, this may mean that a platform deploying a model known to generate abusive content could face liability — even if the harmful content was not created intentionally.

In sum, Brazil combines criminal law, civil liability, and platform regulation to address AI-generated explicit content. While Article 21 of the MCI serves as a powerful tool for protecting individuals from nonconsensual explicit exposure, PL 2338/2023 pushes the conversation further by embedding proactive obligations and safeguards into the AI development pipeline. As jurisprudence evolves, we are likely to see these frameworks tested — and potentially expanded — in response to the unique risks posed by synthetic media.

2.4.5. CSAM Takedowns, New Legislation on Protecting Children Online and Its Impacts on Generative AI Tools

In early August 2025, a 50 minute YouTube video by the Brazilian creator Felca (Felipe Bressanim Pereira) set off a national reckoning about the "adultization" of minors on social media. The video marshaled examples to argue that platform incentives and recommendation systems helped normalize sexualized depictions of minors and facilitated predatory behaviors. By August 12, Felca's video had motivated 32 new bills on child protection online in the National Congress, underscoring how a single piece of user generated content can trigger sweeping regulatory momentum.

Against that backdrop, the National Congress approved PL 2628/2022, nicknamed "ECA Digital" for its alignment with the Child and Adolescent Statute (ECA). The approved version prohibits monetizing or boosting content that erotizes minors and creates structured processes for removal upon notification by

23 Bill No. 2338/2023, art. 25 (2024).

restricted actors (victims/their representatives, the Public Prosecutor's Office, or accredited child rights entities), with contestation and due process mechanisms for users.[24]

While PL 2628/2022 does not create a bespoke regime for foundation models or mandate deepfake labeling, its definitions and obligations cover any "product or service of information technology" that is directed to, or is likely to be accessed by, minors. In practice this includes chatbots, creative AI apps, recommendation systems, and AI augmented features inside games, social networks, and app stores. The bill makes that plain by (1) imposing age gating and parental supervision duties across app stores and operating system layers; (2) prohibiting profiling for targeted advertising to minors (including by techniques such as emotional analysis or AR/VR); and (3) requiring that "tools of artificial intelligence" undergo regular review with expert participation to assure safe use by children and adolescents — language that directly captures generative AI features shipped inside consumer services.

Generative AI providers that are "likely to be accessed" by minors must implement age appropriate design and default high protections (e.g., easy to use controls, the ability to disable personalized recommendations, anticompulsion user experience, or UX), and they must be able to demonstrate risk assessment and mitigation for child users — obligations that naturally extend to prompt based generation features and content filters.

2.5. Hate Speech

Brazilian law prohibits hate speech through a combination of constitutional protections, criminal sanctions, and civil liability mechanisms. However, the application of these rules to AI-generated hate speech presents new challenges for enforcement, responsibility, and rights balancing — especially when synthetic content mimics human expression without having a clear author.

2.5.1. Legal Framework

As previously mentioned, the Federal Constitution guarantees freedom of expression. Brazil is also a party to the International Convention on the Elimination of All Forms of Racial Discrimination, which has informed national legislation against discriminatory speech.

The Penal Code criminalizes *injúria racial,* or racial slurs. Additional provisions from Law No. 7.716/1989 criminalize the incitement of discrimination or prejudice based on race, ethnicity, religion, or national origin. And in 2023, the Federal Supreme Court (STF) ruled that hate speech against the LGBTQIA+ community must receive the same constitutional treatment as racist speech, further expanding the reach of criminal liability in this area.[25]

2.5.2. AI and the Problem of "Non-Human Speakers"

These statutes assume that a human subject authored or disseminated the harmful speech. But generative AI disrupts this logic. When an LLM outputs discriminatory or hateful text, is it "speech"? And if so, who is the speaker?

24 PL 2628/22, "Projeto aprovado proíbe provedores de monetizar conteúdos que viole direitos da criança," House of Representatives, August 21, 2025 (Braz.), https://www.camara.leg.br/noticias/1191259-projeto-aprovado-proibe-provedores-de-monetizar-conteudo-que-viole-direitos-da-crianca
25 STF, *Mandado de Injunção* (MI) No. 4733, Justice Edson Fachin, August 22, 2023 (Braz.).

Current jurisprudence does not offer a clear answer. However, from a regulatory standpoint, there is growing consensus in Brazil that AI-generated content must be traceable and governed by human responsibility, especially in contexts that implicate fundamental rights.

PL 2338/2023 addresses the issue of discrimination and hate speech in several ways. First, it establishes the promotion of equality, pluralism, and nondiscrimination as foundational principles of AI governance. It also defines "abusive and illicit discrimination" and includes this as a factor in identifying high-risk applications.[26]

AI systems that generate, distribute, or amplify discriminatory or hateful content may be classified as high risk. If so, they are subject to governance obligations such as algorithmic impact assessments; documentation of bias-mitigation efforts; transparency and human oversight; and reversibility and redress mechanisms for affected individuals.

These requirements apply not only to systems that explicitly produce hate speech but also to recommender algorithms or moderation tools that might suppress or amplify certain viewpoints in ways that disadvantage protected groups.

2.5.3. Liability and the Role of Risk-Based Regulation

Brazil's general tort law and consumer protection regimes allow for strict liability in cases where harm arises from risky activities or defective services. As discussed in the section on defamation, the Civil Code (Article 927 specifically) and the Consumer Protection Code provide a strong basis for holding developers and deployers accountable, even without fault, especially when the AI system is known to generate biased or hateful results.

The bill also allows courts to shift the burden of proof, which is particularly important in discrimination cases where victims may not have access to model data, training documentation, or output logs. This procedural innovation represents a significant evolution in how hate speech liability could be litigated in the AI era.

2.5.4. Online Platforms and Moderation

In platform environments, hate speech is typically addressed through content moderation systems. Under the 2014 MCI, platforms were not liable for third-party content unless they had failed to comply with a judicial takedown order. However, if the hate speech is generated by the platform's own AI model, this protection may not apply.

Moreover, in the last round of discussions in the Federal Senate, a provision was added to PL 2338/2023 to restrict its enforcement on automated content moderation systems. Article 77 provides that "the regulation of aspects related to the circulation of online content that may affect freedom of expression, including the use of AI for content moderation and recommendation, may only be carried out through specific legislation."[27]

26 Bill No. 2338/2023, art. 4, XI–XII; art. 15, II (2024).
27 Bill No. 2338/2023, art. 77 (2024).

2.5.5. Practical and Enforcement Challenges

Despite the legal tools available, enforcement of hate speech laws, especially in the context of AI, remains difficult. Some challenges include (1) opacity of training data and model behavior, which may embed or replicate societal biases; (2) difficulties in detecting AI-generated hate speech, especially when phrased in coded or indirect ways; and (3) jurisdictional limitations, as AI models may be developed abroad and accessed through global platforms.

Nevertheless, Brazil's evolving framework, anchored as it is in risk-based regulation, civil liability, and antidiscrimination principles, provides a growing foundation for addressing these issues.

2.6. Election and Political Content

Disinformation has become a central concern in Brazil's efforts to regulate both digital platforms and artificial intelligence. While Brazil recognizes freedom of expression as a constitutional right, the manipulation of public discourse through synthetic or deceptive content, especially during elections, has led to a growing number of legislative and regulatory interventions. Generative AI has intensified these challenges by enabling the scalable creation of deepfakes, automated political spam, and other synthetic content.

2.6.1. Electoral Regulation: Resolution No. 23.732/2024

In February 2024, Brazil's Superior Electoral Court (TSE) issued Resolution No. 23.732, establishing a framework to regulate the use of AI and to combat disinformation during the 2024 electoral cycle. The resolution includes several landmark provisions:

- **Prohibition of AI to spread false content**: Candidates and political parties are expressly prohibited from using generative AI to create or disseminate misleading content that distorts facts, manipulates audiovisual materials, or impersonates people.
- **Obligatory labeling**: Any content produced or altered by AI must explicitly disclose its synthetic nature, either through visible labeling in the content itself or via metadata.
- **Platform obligations**: Internet application providers are required to implement measures to detect and limit the spread of manipulated or illicit content, especially if it threatens the integrity of the electoral process. This includes removal obligations once notified by the Electoral Justice system.

The resolution also introduced a prohibition on the use of avatars or virtual characters impersonating candidates during the electoral campaign period. These decisions reflect a growing concern with the rise of synthetic media and its potential to deceive voters, especially in a context of low media literacy. By banning AI-generated avatars, the TSE has aimed to preempt confusion between real and simulated personas; these avatars, while visually compelling, may be powered by LLMs capable of producing campaign messages without supervision or proper controls. The court's rationale considers the limited public understanding of how such interfaces operate, treating these avatars not merely as stylistic choices but as potentially manipulative tools in the voter-candidate relationship.

The blanket prohibition on avatars raises important questions about freedom of expression in electoral contexts. While designed to curb manipulation and disinformation, the ban may also restrict innovative

and accessible forms of political engagement. Campaigns targeting younger audiences or digital-native communities might find in avatars an effective and culturally resonant medium. Moreover, if accompanied by transparency and clear disclaimers, the use of AI-generated spokespeople could enhance, rather than hinder, democratic participation. The court's decision in this situation underscores the tension between safeguarding the electorate and preserving expressive experimentation in campaign strategies — an unresolved issue in the broader debate about regulating generative AI in political communication.

This resolution therefore joins other regulatory efforts globally that address AI and electoral integrity, anticipating not just the manipulation of public opinion but also the difficulty in detecting synthetic political content in real time.[28]

2.6.2. Content Provenance, Transparency, and Labeling

The AI Bill (PL 2338/2023) does not create a bespoke regime for electoral AI use, but its integrity provisions are designed to align with sector-specific regulation like the TSE resolution just discussed.

Resolution No. 23.732 and PL 2338/2023 converge around the principle that transparency and labeling are essential to managing the risks posed by generative AI. They require developers and deployers to inform users when content has been artificially generated or altered.

This focus on disclosure as a safeguard represents a shift away from traditional reactive models (e.g., removal after notice) and toward preventive regulation, with the aim of reducing the virality and credibility of misleading AI-generated content.

2.6.3. Freedom of Expression Concerns

While these provisions seek to protect democratic processes in Brazil, they also raise freedom of expression concerns. The requirement to label AI-generated content must be designed carefully to avoid chilling legitimate uses of satire, parody, or artistic political commentary. Similarly, enforcement mechanisms must ensure due process and guard against over-removal or preemptive censorship of critical voices under the guise of combating disinformation.

Encouragingly, a recent amendment to PL 2338/2023 revises the concept of "information integrity" to explicitly prevent its misuse as a basis for censorship, emphasizing that this notion should be instrumental in promoting, not limiting, expressive freedom.[29]

28 Catherine Régis, Florian Martin-Bariteau, Okechukwu (Jake) Effoduh, Juan David Gutiérrez, Gina Neff, Carlos Affonso Souza, and Célia Zolynski, "AI in the Ballot Box: Four Actions to Safeguard Election Integrity and Uphold Democracy," Toronto Metropolitan University, February 10, 2025, https://doi.org/10.32920/28382087.v1.

29 Article 2, XV of the AI Bill states that the development, implementation and use of AI systems in Brazil have information integrity among its fundaments, "through the protection and promotion of trust, precision and consistency of information for the strengthening of freedom of expression, access to information and other fundamental rights". Article 4, XXII defines information integrity as "the result of an information ecosystem that makes trusted, diverse and precise information and knowledge accessible in a timely manner to promote freedom of expression." Bill No. 2338/2023, arts. 2, XV; 4, XXII (2024).

2.7. Copyright

Copyright law in Brazil, grounded in Law No. 9.610/1998, provides protection for original literary, artistic, and scientific works. This regime extends to software under Law No. 9.609/1998, which is often applied by analogy to AI systems. However, the Brazilian legal system faces significant challenges in applying these frameworks to generative AI, particularly regarding the use of protected works in training data and the ownership of AI-generated outputs.

2.7.1. Use of Copyrighted Materials in Training Data

One of the most pressing questions in the generative AI context is whether using copyrighted material to train AI models constitutes infringement. Currently, Brazilian law does not expressly regulate this practice. While software and databases may be protected, the law does not provide clarity on text and data mining (TDM) for training purposes. Consequently, developers operate in a legal gray zone, often relying on public domain content or open-licensed works to mitigate potential liability. Using copyrighted material without consent or unless clearly grounded in an exception or limitation could lead to litigation. The most famous Brazilian newspaper company, Folha de São Paulo, has sued OpenAI for copyright infringement, claiming that its articles and other proprietary content have been used to train ChatGPT without authorization. Folha requested the "destruction of GPT models that have incorporated such content."[30]

The uncertainty around TDM is partly addressed by PL 2338/2023. The bill establishes obligations for developers of general-purpose and generative AI systems, including the requirement to process only data collected and treated in conformity with legal standards, especially data protected by copyright, and to publish a summary of the datasets used in training.

These provisions are intended to increase transparency and accountability, but they raise practical and technical challenges. Publishing summaries of datasets can be especially complex in the context of large-scale, opaque training pipelines that scrape vast quantities of data from the internet. Critics have pointed out that compliance with such rules may be unfeasible without a harmonized international approach or clearer technical standards.[31]

Furthermore, PL 2338/2023 introduces a narrow exception for the use of copyrighted content in AI training when conducted by public-interest entities (e.g., research institutions, libraries, archives). The exception is contingent on noncommercial use and proper access rights. It does not extend to commercial developers, who must ensure their training data is lawfully obtained — either through licensing, use of open access materials, or drawing on data that is not protected.

The lack of a fair use doctrine comparable to that of the United States also makes it more difficult to justify expansive training datasets under Brazilian law. While the Constitution does recognize the social function of intellectual property, this principle has not been translated into exceptions for AI training under the current law.

30 Patricia Campo Mello, "Folha entra com ação contra OpenAI por concorrência desleal e violação de direitos autorais," *Folha de S. Paulo*, August 22, 2025, https://www1.folha.uol.com.br/mercado/2025/08/folha-entra-com-acao-contra-openai-por-concorrencia-desleal-e-violacao-de-direitos-autorais.shtml.
31 Pedro Henrique Ramos, Julia de Albuquerque Barreto, Marina Garrote, and Stephanie Mathias de Souza, "Remuneração por direitos autorais em IA: Limites e desafios de implementação," Policy Briefs Reglab, no. 3 (May 20, 2025) (Braz.).

2.7.2. Authorship and Ownership of AI-Generated Content

Brazilian copyright law is explicit in requiring human authorship. The Copyright Act defines the author as a natural person.[32] Thus, unless a human provides the creative input — such as crafting prompts or curating outputs — content generated by AI is considered unprotected and in the public domain.

In practice, this limitation affects not only the potential to claim exclusive rights but also the ability to enforce ownership or prevent misuse of AI-generated content. The legal status of AI outputs depends heavily on how courts interpret the level of human creativity involved in their production.

PL 2338/2023 does not confer copyright protection on AI systems or their outputs; it reinforces that developers of generative models must implement risk assessments and be transparent about training and output risks, including potential infringement of third-party rights. This indirectly links to copyright concerns by increasing the compliance burden on developers to preemptively identify and mitigate legal risks.

2.7.3. Commentary on Legislative Adequacy

The Brazilian approach, as reflected in PL 2338/2023, aligns in part with the risk-based framework of the EU's AI Act, but it goes further in mandating transparency for training data. Although well intentioned, this requirement may be impractical for commercial models that rely on large, opaque datasets scraped from across the web.

Moreover, the bill's limited exception for TDM fails to resolve the broader tension between innovation and rights-holder interests. The lack of safe harbors or expansive exceptions similar to "fair use" may hinder domestic AI development, particularly for small enterprises and open-source initiatives.

In the absence of further legislative clarification or judicial precedent, the copyright landscape for generative AI in Brazil remains uncertain and potentially risky for developers. The combination of strict authorship rules, dataset transparency obligations, and exceptions with disputed interpretation makes Brazil's current regime conservative compared to those of other jurisdictions, such as Japan and the United States.

2.8. Measures Empowering Freedom of Expression

While much of the legislative focus on AI in Brazil has centered on risks, prohibitions, and liability, the country has also seen notable efforts — both within and outside of government — to leverage AI in ways that expand expressive freedoms, improve access to information, and democratize participation in the digital public sphere.

2.8.1. Legal and Policy Frameworks

Brazil's PL 2338/2023 enshrines freedom of expression as a guiding principle of AI governance. The bill also incorporates the promotion of informational integrity and pluralism as foundational values, reinforcing the idea that AI regulation should support, rather than restrict, the free flow of ideas in a democratic society.

32 Law No. 9.610, Copyright Law, art. 11, February 19, 1998 (Braz.).

The bill also encourages multi-stakeholder governance, promoting collaboration among civil society, academia, and regulators to ensure that human rights — including freedom of expression — are preserved in AI design and deployment.[33]

2.8.2. Language Inclusion and Regional Representation

Another dimension of empowerment in the legislation relates to linguistic and geographic inclusivity. The vast majority of foundation models are trained predominantly on English-language data, which can marginalize Portuguese speakers and even more so users of underrepresented regional dialects or indigenous languages in Brazil.

Even though no specific legal mandate exists for language diversity in the AI landscape, PL 2338/2023 calls for the promotion of innovation ecosystems that reflect local and regional realities.[34] This creates a policy opening for public funding and research priorities to support the development of Portuguese-based and Brazil-centered models — especially those reflecting the linguistic, cultural, and racial diversity of the population.

Brazil's academic institutions have also promoted open models and public datasets that can be fine-tuned for local contexts. Public universities and research centers were key contributors to the free and open software movement in the recent decades. These efforts help decentralize AI infrastructure and ensure broader participation in the development of generative tools.

2.8.3. Accessibility and Vulnerable Populations

The AI Bill requires AI systems used with vulnerable populations — including children, the elderly, and people with disabilities — to be developed and implemented in a way that ensures clear, age-appropriate, and cognitively accessible communication.[35] This move helps make generative AI tools usable by broader segments of the population and supports the inclusion of such groups in digital discourse.

More broadly, PL 2338/2023 promotes accessibility through its principles of nondiscrimination, human supervision, and explainability (Articles 2 and 6–7). These provisions aim to prevent AI from becoming a tool of exclusion or gatekeeping in education, employment, or civic engagement.

33 Bill No. 2338/2023, art. 68 (2024).
34 Bill No. 2338/2023, art. 69, III and V (2024).
35 Bill No. 2338/2023, art. 5, §2º (2024).

2.9. Miscellaneous

2.9.1. AI and the Right to Be Forgotten

One of the most challenging intersections of AI, speech, and privacy in Brazil involves the right to be forgotten.[36] Although the Supreme Court ruled in 2021) that this right is not compatible with the constitutional protection of free speech and the right to information,[37] there is room for new debates to emerge concerning the use of generative AI systems under a data protection lens.

This debate becomes particularly acute in systems trained on massive public data, where an AI system may resurface stigmatizing or outdated personal information that has not been readily available. The Brazilian General Data Protection Law (LGPD) provides data subjects with rights to erasure, rectification, and review of automated decisions regarding their personal data online, but these safeguards remain imprecise in the context of AI training and inference, where personal data may be embedded at scale. The complexity of mechanisms for post-training data purging creates a situation in which once a model is trained, retroactive enforcement of data subject rights becomes technically and legally challenging.

2.9.2. Data Protection vs. Freedom of Expression in the Context of Generative AI

Brazil has provided a recent and highly illustrative development on a possible clash between data protection concerns and expressive freedoms when it comes to the training and deployment of generative AI applications. The country's National Data Protection Authority (ANPD) issued an injunction to suspend the rollout of Meta AI in Brazil, based on concerns that the system would process public user content from Instagram and Facebook without adequate legal basis under the LGPD. The ANPD cited the absence of transparent consent, the lack of data minimization, and the potential misuse of user content for AI training purposes, particularly where individuals were not clearly informed or empowered to opt out.[38]

This intervention — though grounded in legitimate privacy concerns and currently revoked — reveals a growing constitutional tension between data protection and freedom of expression. On one hand, protecting individuals' control over their personal data is essential in an age of pervasive AI. On the other, restricting access to public data for training or analysis may inadvertently curtail lawful expression, limit media innovation, or chill the reuse of public discourse in transformative or critical ways.

The Brazilian Constitution enshrines freedom of expression, communication, and access to information, alongside the right to privacy and data protection. As AI models increasingly sit at the intersection of these rights — drawing on public expressions to generate new outputs — legal clarity is urgently needed to ensure data protection enforcement does not unintentionally suppress expressive freedom, and vice versa.

---

36 The right to be forgotten (RTBF), when applied to the Internet and digital media, often refers to an individual's ability to request the removal of personal information from search engines or online platforms when such data is outdated, irrelevant, or disproportionately harmful to their privacy. This right emerged prominently in Europe, crystallized through the 2014 Google Spain case before the Court of Justice of the European Union, and later codified within the General Data Protection Regulation (GDPR). In the European debate, RTBF is framed as a crucial extension of data protection rights, balancing individual privacy with freedom of expression and the public's right to information. In Latin America, however, the adaptation of RTBF principles faces significant challenges. While countries such as Colombia and Brazil have engaged in debates and even judicial rulings involving de-indexation of online information, there are deep concerns about how this right might interact with regional histories of censorship and authoritarianism. For example, critics argue that the RTBF could serve as a tool to obscure the historical record, limiting access to information about public officials or past state abuses. This tension makes the RTBF debate in Latin America uniquely complex: it is not only about privacy and data protection, but also about the collective right to truth and memory. See Edoardo Bertoni, "The Right To Be Forgotten: An Insult to Latin American History", *HuffPost*, September 24, 2014, https://www.huffpost.com/entry/the-right-to-be-forgotten_b_5870664.

37 In 2021, the Brazilian Supreme Federal Court (STF) decided the *Aída Curi* case, which arose from the family of Aída Curi, murdered in Rio de Janeiro in 1958, seeking compensation for the rebroadcast of a television program recounting her story. The family argued that the renewed exposure violated her dignity and invoked a supposed "right to be forgotten" to prevent media outlets from revisiting the case decades later. The STF, however, held that such a right is not compatible with the Federal Constitution, stressing that freedom of expression and the right to information prevail when dealing with truthful, lawfully published facts of historical or journalistic relevance. While the Court acknowledged that individuals may seek remedies against abuses such as defamatory, false, or excessively sensationalist reporting, it rejected the recognition of a generalized right to be forgotten. (STF, Extraordinary Appeal No 1.010.606/RJ, Justice Dias Toffoli, February 11, 2021 (Braz.).

38 Malu Cursino, "Brazil Blocks Meta from Using Social Media Posts to Train AI," *BBC News*, July 3, 2024, https://www.bbc.com/news/articles/c729113nvwvo.

This episode with Meta AI in Brazil also demonstrates the growing assertiveness of the country's data protection authority, ANPD, and its potential to shape not only AI compliance but also the boundaries of lawful data use for expressive purposes. Future jurisprudence needs to reconcile these overlapping domains, ideally through a lens that recognizes both the informational autonomy of individuals and the democratic importance of robust public discourse, including discourse through AI-mediated expression.

2.9.3. AI and the Judiciary

Another emerging area in the intersection of AI models and legislation relates to the use of generative AI in the judicial system itself. Some Brazilian courts have begun experimenting with AI tools to partially draft decisions or assist in legal reasoning. While these initiatives are driven by efficiency goals, they raise concerns about transparency, accountability, and access to legal reasoning.

If court decisions incorporate language generated by AI, litigants must have the right to understand how that content was produced and whether it involved undisclosed biases. PL 2338/2023 addresses algorithmic decision-making in the public sector by requiring human oversight, explainability, and safeguards for due process,[39] but practical implementation of these remains uncertain. Ensuring that expressive rights are preserved within the legal process, particularly for vulnerable or unrepresented litigants, is a priority to keep sight of.

39 Bill No. 2338/2023, arts. 39-40 (2024).

3. Conclusion

Brazil's experience with generative AI regulation offers a nuanced and instructive example of how emerging technologies intersect with long-standing commitments to freedom of expression and democratic values. The country's legal and institutional framework is marked by strong constitutional protections for expressive freedom, an active judiciary, and increasingly sophisticated data protection and digital governance regimes. This foundation has enabled Brazil to respond quickly to the new challenges posed by synthetic media, deepfakes, and automated content generation.

The approval of PL 2338/2023 in the Senate marks a turning point in this process. As one of the most comprehensive AI-specific legislations currently under debate in Latin America, it brings to the fore a rights-based and risk-assessment approach that situates freedom of expression as a value to be protected from AI-related harm as well as a guiding principle in AI governance. It introduces innovative mechanisms — including risk-based classification, disclosure obligations, and a reverse burden of proof — that, while not radically departing from existing liability rules, create a platform for future judicial and regulatory evolution.

Despite all of this, PL 2338/2023 has been criticized for its vague definitions and potential chilling effects on freedom of expression. One of the most contentious aspects of the AI Bill is the broad and ambiguous categorization of "high-risk" AI systems. Another concern is its excessive reliance on regulatory discretion and the possibility of politically motivated enforcement, particularly in contexts involving speech-related technologies. Additionally, the labeling requirements for synthetic content, although meant to foster transparency, may not adequately distinguish between malicious deepfakes and legitimate uses of AI such as parody, art, or activism, thereby risking overreach into constitutionally protected expression.

The challenges ahead are real. Tensions between data protection and expressive use of public information, as seen in the ANPD's temporary suspension of Meta AI, illustrate the difficulty of balancing informational autonomy with the public's right to speak, remix, and critique.

The country's courts are now central actors in this unfolding story. With rulings on the new intermediaries' liability regime and a growing docket of cases involving platform governance and electoral disinformation, the Brazilian judiciary will help shape the thresholds for platform responsibility, user rights, and the legal treatment of AI-generated content.

Brazil's regulatory direction also reveals a strategic opportunity: to expand access to AI as a tool for creation, participation, and public engagement. From support for local language models and open-source development to civil society's advocacy for digital inclusion, Brazil's ecosystem contains a range of ingredients for a more equitable and pluralistic AI future.

Ultimately, the Brazilian approach reflects the complexities of governing AI in a democratic society marked by inequality, innovation, and legal ambition. Whether Brazil succeeds in balancing expressive freedom with rights to dignity, privacy, and democratic integrity depends on the continued interaction among principled legislation, proactive enforcement, and constitutional interpretation. In that process, Brazil has the potential not only to govern AI responsibly at home but also to help set the tone for AI governance across the region and beyond.

THE FUTURE OF FREE SPEECH

OCTOBER 2025

www.futurefreespeech.org