

Lancieri, Filippo Maria; Pereira Neto, Caio Mário da Silva; Karolczak, Rodrigo Moura; De Assis, Barbara Marchiori

Working Paper

Adjudicating fake news

New Working Paper Series, No. 372

Provided in Cooperation with:

George J. Stigler Center for the Study of the Economy and the State, The University of Chicago Booth School of Business

Suggested Citation: Lancieri, Filippo Maria; Pereira Neto, Caio Mário da Silva; Karolczak, Rodrigo Moura; De Assis, Barbara Marchiori (2025) : Adjudicating fake news, New Working Paper Series, No. 372, University of Chicago Booth School of Business, Stigler Center for the Study of the Economy and the State, Chicago, IL

This Version is available at:

<https://hdl.handle.net/10419/336739>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Stigler Center

for the Study of the Economy and the State

ADJUDICATING FAKE NEWS

Filippo Lancieri
Georgetown University

Caio Mario da Silva Pereira Neto
FGV Sao Paulo Law School

Rodrigo Moura Karolczak
University of Illinois Chicago

Barbara Marchiori de Assis
Cornell University

September 2025
New Working Paper Series No. #372

Abstract:

On July 30, 2025, President Trump imposed 50% tariffs on imports from Brazil and sanctioned a sitting Brazilian Supreme Court Justice, both partially because of Brazil's online content moderation decisions. This is an extreme, but not an isolated event: worldwide, legislators and regulators struggle to craft public policies that address problems of disinformation and online harassment while protecting the freedom of expression—leading to increasing international confrontations. One key question in content moderation is content adjudication—or who is responsible for deciding what type of speech violates the law and should be taken down. This article presents the results of a six-year, large empirical and qualitative project on the adjudication of fake news disputes by Brazilian Courts from 2018 onwards. It examines what led Brazilian judges to order the takedown of online content, which social networks and types of content were

most affected by judicial decisions, and whether there is evidence that incumbent politicians abused the system, among other factors. It also critically analyzes the evolution of this novel court-driven content moderation regime—one in which Courts play an increasingly active role in policing online discourse—with significant implications for the Brazilian information ecosystem, democratic institutions, and judicial reputation. Ultimately, the Brazilian experience teaches/reinforces five lessons to jurisdictions rethinking their online content moderation regulatory regimes: (i) content moderation systems must articulate clear end goals to work properly; (ii) experimentation in content moderation regulation is possible and desirable, allowing for incremental learning and adaptation; (iii) a content moderation system must have strict protection against the slippery slope that may lead it to censorship and arbitrariness; (iv) content-based moderation systems become part of the information environment—claims about neutrality in adjudication cannot, by themselves, support long-term systemic legitimacy; (v) increasing regulatory fragmentation impose new urgency on the development of international guidelines for limits on the extraterritoriality of online decisions.

Stigler Center for the Study of the Economy and the State
University of Chicago Booth School of Business
5807 S Woodlawn Ave
Chicago, IL 60637

Adjudicating Fake News

Filippo Lancieri

Caio Mario da Silva Pereira Neto

Rodrigo Moura Karolczak

Barbara Marchiori de Assis

Abstract:

On July 30, 2025, President Trump imposed 50% tariffs on imports from Brazil and sanctioned a sitting Brazilian Supreme Court Justice, both partially because of Brazil's online content moderation decisions. This is an extreme, but not an isolated event: worldwide, legislators and regulators struggle to craft public policies that address problems of disinformation and online harassment while protecting the freedom of expression—leading to increasing international confrontations. One key question in content moderation is content adjudication—or who is responsible for deciding what type of speech violates the law and should be taken down. This article presents the results of a six-year, large empirical and qualitative project on the adjudication of fake news disputes by Brazilian Courts from 2018 onwards. It examines what led Brazilian judges to order the takedown of online content, which social networks and types of content were most affected by judicial decisions, and whether there is evidence that incumbent politicians abused the system, among other factors. It also critically analyzes the evolution of this novel court-driven content moderation regime—one in which Courts play an increasingly active role in policing online discourse—with significant implications for the Brazilian information ecosystem, democratic institutions, and judicial reputation. Ultimately, the Brazilian experience teaches/reinforces five lessons to jurisdictions rethinking their online content moderation regulatory regimes: (i) content moderation systems must articulate clear end goals to work properly; (ii) experimentation in content moderation regulation is possible and desirable, allowing for incremental learning and adaptation; (iii) a content moderation system must have strict protection against the slippery slope that may lead it to censorship and arbitrariness; (iv) content-based moderation systems become part of the information environment—claims about neutrality in adjudication cannot, by themselves, support long-term systemic legitimacy; (v) increasing regulatory fragmentation impose new urgency on the development of international guidelines for limits on the extraterritoriality of online decisions.¹

¹ Lancieri: Associate Professor of Law, Georgetown University Law Center; Fellow, The George J. Stigler Center for the Study of the Economy and the State, UChicago Booth;

Table of Contents

I. INTRODUCTION	3
II. ELECTIONS AND THE ROLE OF THE BRAZILIAN JUDICIARY IN TACKLING DISINFORMATION	9
III. HOW DID THE BRAZILIAN SYSTEM PERFORM IN PRACTICE?	12
A. DISINFORMATION ADJUDICATION IN THE 2018 BRAZILIAN ELECTIONS	13
B. DISINFORMATION ADJUDICATION IN THE 2022 BRAZILIAN ELECTIONS	24
C. THE 2022-2025 AFTERMATH: AN INCREASINGLY POWERFUL “CONTENT-MODERATION” COURT	27
IV. ASSESSING THE BRAZILIAN CASE	34
A. FIVE THEORETICAL CRITERIA TO ASSESS THE BRAZILIAN CASE	34
B. APPLYING THESE CRITERIA TO THE BRAZILIAN COURT-BASED CONTENT MODERATION SYSTEM.....	35
C. FIVE MORE GENERAL TAKEAWAYS FOR OTHER JURISDICTIONS.....	37
V. CONCLUSION	43
APPENDIX I: MEASURING ASSOCIATIONS IN THE TAKEDOWN ORDERS FOR THE 2018 ELECTION DECISIONS	44

Pereira Neto: Professor of Law, FGV Direito SP. Coordinator of the Competition, Public Policy, Innovation, and Technology (COMPPIT) Nucleus at FGV Direito SP. Partner at Pereira Neto, Macedo, Rocco;

Karolczak: PhD candidate in Political Sciences, University of Illinois at Chicago;

Marchiori de Assis: Research Fellow at the COMPPIT Nucleus, Senior Manager in Cyber and Strategic Risk at Deloitte Italy, formerly at the IDB and OAS

We thank Stefan Bechtold, Anu Bradford, Eric Talley, Joshua Tucker, William Hubbard, Amit Zac and participants at academic workshops and conferences at Chicago, Columbia, ETH Zurich, FGV Direito Sao Paulo, Georgetown, Harvard, SciencesPo, Tilburg, the University of Zurich, and Yale for feedback on different versions of this research. The coding of the first round of disinformation decisions was carried out by the Research Group on Competition Policy and Regulation of Digital Platforms at FGV Direito SP between 2019 and 2022, with the collaboration of the following members: Antonio Bloch Belizario; Daniel Favoretto Rocha; Esther Simon Seroussi Souccar; Fernanda Mascarenhas Marques; Georges Vicentini El Hajj Moussa; Giulia de Paola; Helena Secaf dos Santos; Pedro Marques Neto; Raíssa Leite de Freitas Paixão; and Vitória Oliveira. Antonio Bloch Belizario, Helena Ambiel and Lucas Bertolo provided valuable research assistance regarding the 2022 elections and beyond. We are very thankful for their support

According to the ASCOLA Declaration of Transparency and Disclosure, we report:

Lancieri: I worked on a CERRE-supported report on access to data and algorithms under the DMA/DSA framework. Pereira Neto: As a partner at Pereira Neto, Macedo, Rocco I advise digital platforms such as Google, Airbnb, and Uber on matters relating to Brazilian antitrust and digital regulations. These companies did not provide financial or material support nor were directly involved in any stage of the drafting of this article. Karolczak: Karolczak was part of a team that received an unconditional grant from Facebook to develop the FGV/CEPI database of disinformation cases during the Brazilian 2018 elections, published in 2019. This article partially relies on that database. Marchiori de Assis: nothing to disclose.

I. Introduction

It was around 6 p.m. on Tuesday, October 8th, 2024, when Twitter/X announced it was “proud” to restart its Brazilian operations². This followed a previous, less broadcast declaration that the company would now fully comply with all Brazilian laws and court orders. It had been over two months since tens of millions of Brazilian users lost access to X’s services due to a Supreme Court ban, following the company’s refusal to remove multiple accounts deemed to spread disinformation in Brazil.

The decision to ban X from the world’s seventh most populous country (215 million people) was as bold as it was controversial, and placed one Brazilian Supreme Court Justice—Mr. Alexandre de Moraes—in direct confrontation with Elon Musk, the world’s richest man. Many praised Brazil’s courage and its capacity to ensure that even the largest digital platforms comply with national laws.³ Others, such as the Committee on the Judiciary of the US House of Representatives, had been denouncing the Brazilian Supreme Court for its blatant censorship of conservative voices.⁴ Indeed, this decision triggered a cascade reaction with unpredictable consequences. A couple of months later, President Trump’s family company, Trump Media, sued Justice Alexandre de Moraes in the United States for damages arising from the alleged violation of US freedom of expression laws.⁵ Then in late July 2025, an Executive Order from President Trump declared a national emergency and ordered the imposition of 50% tariffs on most US imports from Brazil, citing as a partial motivation how “Members of the Government of Brazil have taken actions that interfere with the economy of the United States, infringe the free expression rights of United States persons, violate human rights, and undermine the interest the United States has in protecting its citizens and companies”⁶. Justice Moraes was singled out: “For example, Brazilian Supreme Court Justice Alexandre de Moraes has abused his judicial authority to target political opponents, shield

² See the post at: <https://x.com/GlobalAffairs/status/1843773083783471139>

³ Daniel Carvalho & Simone Iglesias, *Brazil Took on Musk and Won. Now Lula Is Sharing Notes With Europe*, BLOOMBERG.COM, Feb. 3, 2025, <https://www.bloomberg.com/news/articles/2025-02-03/brazil-took-on-musk-and-won-now-lula-is-sharing-notes-with-europe>.

⁴ See House Committee on the Judiciary, *Brazilian Government Forced Censorship on X: New Report Reveals | House Judiciary Committee Republicans*, (2024), <http://judiciary.house.gov/media/press-releases/brazilian-government-forced-censorship-x-new-report-reveals>.

⁵ Joshua Goodman, *Trump Media Company Sues a Brazilian Supreme Court Justice Investigating Bolsonaro*, AP NEWS (Feb. 19, 2025), <https://apnews.com/article/trump-lawsuit-brazil-judge-bolsonaro-0061c2f1ea145e3ce3714aa25f49ba67>.

⁶ President Donald J. Trump, *Executive Order 13808: Addressing Threats to The United States by the Government of Brazil*, (2025), <https://www.whitehouse.gov/presidential-actions/2025/07/addressing-threats-to-the-us/>.

corrupt allies, and suppress dissent, often in coordination with other Brazilian officials. (...) In fact, Justice de Moraes is currently overseeing the Government of Brazil’s criminal prosecution of a United States resident for speech he made on United States soil.”⁷ Justice de Moraes was himself punished under the Magnitsky Act⁸, a U.S. statute passed to punish individuals responsible for grave human rights violations—usually employed to punish warlords, this was the first time the act was used against a sitting Supreme Court Justice of another nation.

Brazilian content moderation decisions have effectively put the country at the center of a trade war with the world’s most powerful nation.

As we explore in detail in this article, this conflict is partially the outgrowth of a system that relies on the Brazilian judiciary to prevent the spread of disinformation during election periods—initially established in response to the 2016 U.S. Presidential Elections that first elected President Trump. Indeed, this is an extreme, yet not an isolated, episode of the growing international rift over boundaries for online speech, and the geopolitical tensions that arise as countries attempt to enforce their laws⁹. For example, the US government has also been increasingly critical of European Union digital regulations for their perceived censorship of what Americans deem lawful speech¹⁰ and alleged exploitation of American companies and interests. Europeans, in response, argue that they are simply enforcing European laws and defending European sovereignty and values¹¹.

The unescapable reality is that the digital revolution placed the internet at the center of all political debates. Twitter/X, Facebook, Instagram, WhatsApp, YouTube, and increasingly TikTok (among others) have quickly become crucial communication platforms to anyone aspiring to occupy an elected position. At the same time, these platforms empowered the voices of previously excluded citizens and activists, changing the dynamics in the public sphere. However, major electoral upsets around the world have also brought “fake news” and content moderation decisions more broadly to the forefront of threats to

⁷ *Id.*

⁸ *Treasury Sanctions Alexandre de Moraes*, U.S. DEPARTMENT OF THE TREASURY (Feb. 8, 2025), <https://home.treasury.gov/news/press-releases/sb0211>.

⁹ For a framework explaining the international game played by countries and companies in tech regulation, see Filippo Lancieri, Laura Edelson & Stefan Bechtold, *AI Regulation: Competition, Arbitrage & Regulatory Capture*, 26 THEOR. INQ. LAW (2025), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5049259.

¹⁰ See post from the US Department of State openly criticizing the EU Digital Services Act as a censorship tool to prevent citizens to criticize their governments, available at <https://x.com/StateDept/status/1947755665520304253>

¹¹ Susan Gravatt, *EU Competition Chief Says Bloc ‘Won’t Compromise Sovereignty’ in Talks with U.S., Criticizes American Big Tech Crackdown*, THE CAPITOL FORUM (June 30, 2025), <https://thecapitolforum.com/eu-competition-chief-bloc-wont-compromise-sovereignty/>.

democracy, a position it continues to hold¹². When addressing this problem, governments and digital platforms face a fundamental regulatory challenge: how to ensure freedom of expression and the spread of ideas, even if they are controversial—a core value in liberal democracies—while preventing disinformation from fully contaminating the public sphere and undermining the possibility of collective deliberation.

Visions on the limits of freedom of expression vary significantly, and so do policy responses. The Czech Republic experimented with Governmental fact-checking.¹³ Germany's hate speech and false news law imposes fines up to EUR 50 million for violations.¹⁴ The UK requires platforms to enforce their own promises to moderate harmful content.¹⁵ France passed a law against information manipulation, creating a legal injunction system to remove false online information during the three months preceding an election.¹⁶ Singapore granted the Government authority to order the removal of statements of fact deemed false or misleading, allowing for fines or imprisonment in cases of violation.¹⁷ The EU initially relied on self-regulation, only to realize its many limitations and enact the Digital Services Act as a leading example of a regulatory regime for online content moderation worldwide.¹⁸ The US, on the other hand, continues to rely on self-regulation at the Federal Level, even if

¹² See *Global Risks 2024: Disinformation Tops Global Risks 2024 as Environmental Threats Intensify*, WORLD ECONOMIC FORUM, <https://www.weforum.org/press/2024/01/global-risks-report-2024-press-release/>.

¹³ Anthony Faiola, *As Cold War Turns to Information War, a New Fake News Police Combats Disinformation*, WASHINGTON POST, Jan. 22, 2017, https://www.washingtonpost.com/world/europe/as-cold-war-turns-to-information-war-a-new-fake-news-police/2017/01/18/9bf49ff6-d80e-11e6-a0e6-d502d6751bc8_story.html.

¹⁴ Anthony Faiola & Stephanie Kirchner, *How Do You Stop Fake News? In Germany, with a Law.*, WASHINGTON POST, Apr. 5, 2017, https://www.washingtonpost.com/world/europe/how-do-you-stop-fake-news-in-germany-with-a-law/2017/04/05/e6834ad6-1a08-11e7-bcc2-7d1a0973e7b2_story.html.

¹⁵ See UK Home Office, *Online Harms White Paper - Executive Summary*, GOV.UK (Apr. 2019), <https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper-executive-summary--2>, Joe Tyler-Todd & John Woodhouse, *Preventing Misinformation and Disinformation in Online Filter Bubbles - UK House of Commons*, (2024), <https://commonslibrary.parliament.uk/research-briefings/cdp-2024-0003/>.

¹⁶ Loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information (Dec. 22, 2018), <https://perma.cc/QH5N-25MC>; Nicolas Boring, 'Government Response to Disinformation on Social Media Platforms: France' (*Library of Congress*, September 2019) <https://www.loc.gov/law/help/social-media-disinformation/france.php>

¹⁷ See Singapore's Singapore's Parliament, *Protection from Online Falsehoods and Manipulation Bill*, (2019), <https://sso.agc.gov.sg/Bills-Supp/10-2019/Published/20190401?DocDate=20190401>.

¹⁸ See European Commission, *The Digital Services Act Package | Shaping Europe's Digital Future*, (July 28, 2025), <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>.; Daphne Keller, *The DSA's Industrial Model for Content Moderation*, VERFASSUNGSBLOG (2022).

States are increasingly active in passing their own targeted content moderation laws.¹⁹

As countries, companies, and societies scramble to find a solution to disinformation through new legal or extralegal requirements, some questions remain largely undiscussed or lack empirical analysis: under what standards should governments intervene? Which authority should be responsible for enforcing these new regulatory standards? And how to design standards that are effective in preventing the spread of online disinformation *and* can be used by authorities to consistently adjudicate disputes in different contexts and across time?²⁰ We are in the middle of a new era of internet and content moderation, but decades of broad safe harbors for internet speech mean we have few data points to learn from.

This article sheds light on these questions by presenting the results of a six-year-long, empirical and qualitative case study of Brazil. As the world's tenth-largest economy, fourth-largest democracy, and a highly connected nation, Brazil has long been at the frontline of the tug-of-war between democratic governments and digital platforms regarding content moderation.²¹ This transformed Brazil into a laboratory for innovative policy interventions, and over the years, the country established a unique court-based system of content moderation and adjudication that is responsible for resolving online disputes related to electoral disinformation. More specifically, during bi-annual election periods (every even year), an army of around 3,000 career State and Federal Judges and 3,000 State and Federal prosecutors is temporarily transferred to

¹⁹ See Alan Z. Rozenshtein, *A New Era of Internet Regulation Is About to Begin*, THE ATLANTIC (July 8, 2025), <https://www.theatlantic.com/ideas/archive/2025/07/supreme-court-pornography-ai-internet/683449/>.

²⁰ A 2019 literature review on disinformation campaigns stressed the urgency for more research on “*the effects of new laws and regulations intended to limit the spread of disinformation*”. See Joshua Tucker et al., *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*, (2018), <https://www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf>, p. 7. While our academic knowledge has progressed ever since (see Daphne Keller & Paddy Leerssen, *Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation*, in SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM (2020).), most of these pressing questions remain largely unanswered.

²¹ Indeed, back in 2012, much before “*fake news*” was a household term, the head of Google in Brazil was arrested for 24 hours after the company refused to comply with court orders requiring the take-down of YouTube accounts deemed to be spreading false information about a candidate for mayor – see *Google Brazil Head Freed; Arrest Stirs Debate - CBS News*, CBS NEWS (Sept. 27, 2012), <https://www.cbsnews.com/news/google-brazil-head-freed-arrest-stirs-debate/>. In 2016, Facebook’s Vice-President for Latin America spent a night in jail after refusing to provide information about messages shared on WhatsApp – see Brad Haynes, *Facebook Exec Jailed in Brazil as Court Seeks WhatsApp Data*, REUTERS, Mar. 1, 2016, <https://www.reuters.com/article/idUSKCN0W34WA/>.

electoral courts to oversee the fairness of the poll, including the veracity of different forms of political advertisement. Since the 2018 Brazilian presidential elections, these judges and prosecutors have also been empowered to adjudicate disputes related to the sharing of online disinformation, granting all candidates, political parties, and prosecutors the ability to sue. Sure enough, “fake news” has become a central topic of Brazilian elections since the 2018 poll that elected President Jair Bolsonaro (deemed the “Trump of the tropics”).²² This makes Brazil one of a kind in this arena for both its geopolitical importance, populational diversity, and novel approach to target fake news that relies on the independence and professionalization of its electoral court-based content moderation system.

One problem with studying Brazil has been the lack of high-quality data that can enable a comprehensive case study mapping how the system evolved over time. We overcome this problem by splitting our analysis into three major time points. First, we employed a combination of automated and manual coding techniques to create a novel, comprehensive database of all 1,500 disinformation cases adjudicated by the Brazilian courts during the 2018 elections. Second, we complement this comprehensive database with a sample of hundreds of key cases adjudicated by Brazilian courts during the 2022 elections, when President Lula beat former President Bolsonaro in a process similar to what happened with Presidents Trump and Biden during the 2016 and 2020 US elections. Finally, we follow the evolution of the Brazilian Supreme Court’s “Anti-Fake News” inquiry, which started before the 2020 elections to protect the fairness of the poll, but continues to this day—it was under this broad inquiry that Justice Moraes ordered the banning of X in Brazil for spreading disinformation, later triggering retaliation by the United States.

Our results paint a complex and nuanced picture of the ups and downsides of the Brazilian experience, offering important lessons for other jurisdictions around the world. In many ways, the Brazilian Court-led model is a step forward in democratic accountability compared to systems that rely primarily on self-regulation by platforms or opaque administrative structures controlled by unelected bureaucrats—it is transparent, independent, and professionalized. It also effectively enabled the country to impose its laws over the largest digital platforms around the world, something that even important jurisdictions such as the European Union have recurrently failed to accomplish.²³ However, the Brazilian system also has significant limitations: (i)

²² Bevin Vincent, *Trumps of the Tropics: Brazil’s Far Right Plots Its Return*, THE NEW YORK TIMES, Aug. 1, 2024, <https://www.nytimes.com/2024/08/01/magazine/brazil-bolsonaro-trump.html>.

²³ About the failure of EU digital interventions, see Filippo Lancieri & Caio Mario Pereira Neto, *Designing Remedies for Digital Markets: The Interplay between Antitrust and*

it faced and continues to face difficulties in developing coherent standards that can be replicated at scale; (ii) its acknowledgment of the need for fast decisions during elections lead to many unsubstantiated rulings that are hard to square together and the fragmented decision-making focused on specific links is insufficient to deal with systematic spread of disinformation by organized groups; and (iii) the increasing centralization of adjudications in the Superior Electoral Court and the Supreme Court restricted parties ability to file appeals and led to an increased politicization of the judiciary, eroding trust in the higher Court as neutral arbiters of disputes.

The Brazilian case also provides five more general lessons to the broader world, which still struggles with difficulties to tackle widespread disinformation: (i) any content moderation system must articulate clear end goals to work properly—for example, focusing in establishing operational standards to be applied at scale by platforms and civil society; (ii) experimentation in content moderation regulation is clearly possible and desirable, allowing for incremental learning and adaptation, disavowing arguments that it will break the internet; (iii) a content moderation system must have strict protection against the slippery slope that may lead it to censorship and arbitrariness; (iv) any content-based moderation system becomes part of the information environment, so that claims about neutrality in adjudication cannot, by themselves, support long-term systemic legitimacy; (v) increasing regulatory fragmentation impose new urgency on the development of international guidelines for limits on the extraterritoriality of online decisions.

To perform this analysis, this article is divided into five parts, the first of which is this introduction. Part II provides a more detailed explanation of the Brazilian model, outlining how the Brazilian court-based content-moderation regime evolved in tandem with the evolution of online disputes around disinformation content. Part III covers the performance of the Brazilian system during the 2018 (Part III.A) and 2022 elections (Part III.B), as well as the ongoing post-2022 Supreme Court anti-fake news investigation that led to the banning and reinstatement of X, as well as the increasing US sanctions against Brazil (Part III.C). Part IV discusses the strengths and weaknesses of the Brazilian model, drawing on a theoretical framework built on the work of Yale Law Professor Jack Balkin to extract five broader lessons applicable to other jurisdictions. A short conclusion follows.

Regulation, 18 J. COMPET. LAW ECON. 613 (2022). (discussing antitrust cases) and ANU BRADFORD, *DIGITAL EMPIRES: THE GLOBAL BATTLE TO REGULATE TECHNOLOGY* (2023). (discussing how jurisdictions must win “vertical battles” with digital platforms more broadly).

II. Elections and the role of the Brazilian judiciary in tackling disinformation

The contemporary Brazilian democracy was reinstated 40 years ago, in 1985, following 20 years of military dictatorship. The country developed a complex and mature electoral system that has produced peaceful transitions of power over the past four decades. Elections take place in October, every even year, on a Sunday. Every four years (e.g., 2014, 2018, 2022, 2026), the country elects a new President, a new federal House of Representatives, and a portion of the Federal Senate,²⁴ new state Governors, and new state Congresses. Two years after (e.g., 2020, 2024, 2026), the country votes on Mayors and municipal legislatures. Elections for President, Governors, and Mayors often feature run-offs, where the two candidates with the most votes face each other when no candidate receives at least 50% of the valid votes in the first round. Since 2015, the Brazilian Supreme Court has prohibited companies from making campaign contributions. As personal donations are also heavily restricted, most Brazilian political parties and campaigns are primarily funded by specially allocated public funds (a practice that scholars have long advocated for in the US).²⁵

As mentioned above, with a population of approximately 215 million people, Brazil is the world's fourth-largest democracy and a highly connected nation. 180 million Brazilians are active social network users (the fifth largest market worldwide),²⁶ and the country is among the largest markets for Facebook, WhatsApp, Instagram, YouTube, TikTok, and Twitter/X. As a result, Brazilian authorities identified "fake news" as a potential threat to the high-stakes 2018 presidential elections and adjusted the election monitoring structures to address it.

Brazil hosts a sophisticated electoral court system that is responsible for organizing fair and safe elections under the 1988 Federal Constitution. Brazil has a career judiciary similar to France. Judges and prosecutors of lower courts are selected based on nationwide exams, where positions are allocated according to the highest test scores (i.e., a merit-based selection system). Judges and prosecutors slowly move up the ranks until they reach appeal courts or leadership positions²⁷. As in the US, Supreme Court Justices are appointed by the President and confirmed by the Senate. Judges and prosecutors can only be

²⁴ Each state is represented by 3 Senators, which serve an 8-year term. Senatorial elections are organized every four years, together with the presidential election, renewing 1/3 or 2/3 of the Senate every other election.

²⁵ See, for example, LAWRENCE LESSIG, *REPUBLIC, LOST: HOW MONEY CORRUPTS CONGRESS-AND A PLAN TO STOP IT* (2011).

²⁶ See *Topic: Social Media Usage in Brazil*, STATISTA, <https://www.statista.com/topics/6949/social-media-usage-in-brazil/>.

²⁷ With some small exceptions for quotas that ensure a minimum level (around 10%) of political appointments in some appeals courts.

fired for gross negligence or criminal convictions, and they enjoy high salaries, significant stability, and independence.

During election periods, typically three months before and after a given poll, an army of around 3,000 State and Federal Judges and 3,000 State and Federal prosecutors is temporarily transferred from general courts to electoral courts to oversee the fairness of the poll. They oversee, for example, candidate registration, electoral advertising, allegations of vote buying and abuse of economic power, and candidates' and political parties' general reports on expenditures, in an effort to minimize corruption. They are divided into three branches: around 2,500 electoral district courts; 27 Electoral Courts of Appeals (*Tribunais Regionais Eleitorais*, or TREs)²⁸; and the Superior Electoral Tribunal (*Tribunal Superior Eleitoral*, or TSE), Brazil's highest electoral court. Career judges and prosecutors are appointed to each of the electoral districts. The TSE is composed of seven justices²⁹. The TREs follow a similar composition, reflecting lower court appointments³⁰. Both the TSE and TREs also appoint auxiliary justices who are responsible for specific matters, whose decisions are appealable to the full court. In summary, it's a complex, independent, and professionalized electoral court system that functions as a hybrid between the US and European systems.

Brazilian Courts have been trying to address problems associated with online disinformation and irregular electoral advertisements for decades. The electoral advertisement monitoring system was initially established to regulate what candidates for office would state to the public in federally funded TV and Radio prime-time slots allocated to political parties during election periods. Since 2009, the Brazilian electoral law has specific provisions: (i) affirming that online platforms must abide by electoral courts' decisions; and (ii) establishing a safe harbor that exempts platforms from direct liability associated with illegal content as long as companies were unaware of the content's illegal nature. As amended in 2017, the law grants the TSE (the higher electoral court) powers to establish "good practices" regarding political advertisement on online platforms. More controversial, in 2017, the TSE used a formal rulemaking process to expand on such powers and establish restrictions on the dissemination of "*facts known to be untrue*". This became a new formal standard for Courts to order the takedown of false online content during elections. The same regulation also

²⁸ One for each of the 26 Brazilian States and another for the Federal District around the Brazilian Capital—which is similar to Washington D.C. in that it is a Federal District, but different in that residents both elect three Senators and have proportional representation in the Brazilian Federal Congress.

²⁹ Three of these are Brazilian Supreme Court Justices, two are Justices at the Superior Court of Justice (STJ) and two civilians appointed by the Brazilian President from a short list provided by the Supreme Court (usually lawyers or law professors). Seven substitute-judges can replace the justices in case of absences.

³⁰ Instead of Supreme Court Justices, TRE members are Appellate Court judges.

affirms that disinformation is not part of voters' general freedom of speech; therefore, it can be abridged through judicial rulings. Judges justified this change by citing the "fake news" debacle following the 2016 US elections that brought President Trump to power, arguing that they were preparing the Brazilian democracy to avoid the same problems around false and misleading online information.

Brazilian law has also expanded over the years to restrict undue attribution of content to third parties and the use of fake profiles during electoral campaigns. Current law criminalizes or imposes fines against anyone who: (i) unduly attributes the authorship of online electoral advertisement to third parties; (ii) directly or indirectly contracts out individuals or groups to offend candidates or a party through online messages or comments, punishing both the contracting party as well as the contracted parties; or (iii) disseminates content through fake online profiles.

Any candidate, political party, electoral coalition, or public prosecutor may file a complaint alleging a violation of Brazilian Electoral Law. Citizens and companies may not file complaints, but may be defendants. Complaints are usually filed before district electoral courts. District-level decisions can then be appealed to TREs and, in restricted circumstances, to the TSE.

Due to the dynamic nature of polls, these decisions are made under expedited procedural rules: candidates have between 24 and 72 hours to complain about offensive content, counting from when they become aware of such content. Judges must rule on the matter within 24 hours, and usually issue preliminary rulings before hearing defendants. Defendants have 48 hours to either challenge the initial complaint or the interim order. Judges then issue a final verdict. Appeals follow a similar expedited time frame: 24 hours to file the appeal, 48 hours to file counterarguments, and 48 hours for the appellate court to rule. In some cases, complaints may be filed directly before TREs/TSE, depending on the claimant and the nature of the claim. This was the case for disinformation proceedings relating to the 2018 and 2022 Brazilian elections: most cases were filed directly before the 27 different TREs and the TSE, which had each previously appointed three auxiliary judges (84 in total) who would be responsible for handling disinformation claims.

Judges have basically five alternatives when deciding a complaint alleging that someone is spreading disinformation in violation of Brazilian laws:

- (i) decline jurisdiction if the matter does not directly impact the electoral process/results;
- (ii) dismiss the complaint when the challenged information is not sharing "*facts known to be untrue*";

- (iii) grant candidates a “right of reply,” through which the candidate may express their views on the matter in the same venue where the original offense took place;³¹
- (iv) issue a takedown order against the infringing content, including orders prohibiting the content from being reposted online; and/or
- (v) impose civil fines, including fines against platforms to ensure compliance.

These can be applied concomitantly, meaning that judges may grant a right of reply, issue a takedown order, and impose a civil fine as the result of a single complaint. In their decisions, Judges are expressly required to balance out the potential restriction of freedom of expression rights vis-à-vis the potential harm to isonomy between candidates and the fairness of the electoral process. Importantly, the 2017 regulation required that all complaints against online content identify the specific URL linking to that content, meaning that judges initially had to order the takedown of specific URLs. In 2022, as explained below, TSE expanded its powers to also remove content in bulk whenever it was deemed “identical to other previously illegal content”. In exceptional situations, judges may also order the takedown of the entire hosting platform if the platform repeatedly fails to comply with Court rulings. Additionally, all electoral takedown orders are interim in nature and are automatically revoked once the elections are over. This system operates in parallel with civil litigation; parties may file civil lawsuits to require the permanent removal of the content and request compensation for any damages through a regular tort case at any time.

III. How Did the Brazilian System Perform in Practice?

This unique Brazilian system has many theoretical virtues and limitations—a topic we return to in Section IV below. To engage in a critical analysis, however, one must first understand how the system performed over time—the goal of this Part III. We focus on three periods. **Part III.a** builds on a novel database of online disinformation cases to conduct a *comprehensive analysis* of the work of Brazilian Courts in trying to prevent the spread of disinformation during the 2018 Brazilian general elections³²—following the implementation of the system in response to the controversies surrounding the

³¹ For example, political party J is allocated a 10-minute electoral advertisement slot during prime time Brazilian television. During that time, candidate for office A makes a statement against candidate B deemed to be untrue. The Electoral Court, then, can grant candidate B a right of response worth X minutes to be broadcasted during the next 10 minutes that party J holds in that same prime slot in the future.

³² Which elected a new President, State governors, federal and State Congressmen and Federal Senators

2016 US elections. **Part III.b.** covers the 2022 Brazilian Federal/Presidential elections, providing a more restricted analysis of hundreds of key cases during that poll. **Part III.C.** starts with the aftermath of the 2022 elections and goes until July 2025. It covers the increasing centralization of content moderation decisions in the Brazilian Supreme Court and Justice Alexandre de Moraes, the dispute that led to the banning of X from Brazil, and the consolidation of the Brazilian Supreme Court as a key actor in increasingly polarized national political debates—with significant impacts on public perceptions around the politicization and independence of the Court itself.

a. Disinformation Adjudication in the 2018 Brazilian Elections

Following the 2016 US presidential election, fake news was also a central topic in the 2018 Brazilian elections, won by former President Jair Bolsonaro with roughly 55% of the valid votes. Many candidates accused each other of spreading fake news throughout the elections, including a major accusation that Bolsonaro was relying on a professionalized disinformation sharing network to push his political messaging, which surfaced in Brazilian and international media a week before the runoff vote between him and the left-wing Workers' Party candidate, Fernando Haddad.³³ Some international and national commentators allege that the widespread sharing of disinformation (and, more colloquially, “fake news”) shaped the outcome of that election,³⁴ though others attribute Bolsonaro's landslide win to the Workers' Party's multiple corruption scandals and the fielding of a candidate with limited political appeal.³⁵

The central role disinformation and fake news played in the election itself increases interest in understanding the performance of the system created by the Brazilian judiciary to tackle disinformation. This section builds on a large, multi-year effort to identify and manually analyze all decisions issued by Brazilian Courts on the topic over the 2018 election period. More specifically, Brazil has no private or public, comprehensive, searchable case law database of

³³ Tom Phillips, *Bolsonaro Business Backers Accused of Illegal Whatsapp Fake News Campaign*, THE GUARDIAN, Oct. 18, 2018, <https://www.theguardian.com/world/2018/oct/18/brazil-jair-bolsonaro-whatsapp-fake-news-campaign>.

³⁴ Christopher Harden, *Brazil Fell for Fake News: What to Do About It Now?*, WILSON CENTER (Feb. 21, 2019), <https://www.wilsoncenter.org/blog-post/brazil-fell-for-fake-news-what-to-do-about-it-now>.

³⁵ Ricardo Boechat, *A Vitória de Bolsonaro Não é Fruto Das Fake News*, (2018), https://www.youtube.com/watch?v=ym8t7T_c9cw.

all decisions issued by electoral Courts.³⁶ To overcome this problem, FGV-CEPI developed an automated web scraper that downloaded all judge-issued documents available in the Brazilian nationwide electronic procedure system (PJe). The scraper then used a series of regular expression (RegEx) codes and manual review to filter online disinformation cases from the roughly one hundred thousand cases identified and downloaded in the process³⁷. The result was a comprehensive, machine-readable electoral case database for the 2018 Brazilian Elections, containing all 1492 online disinformation cases identified in the process and their respective metadata.³⁸ At the Competition, Public Policy, Innovation, and Technology Nucleus at FGV Direito SP (FGV-COMPPIT), we trained a group of 10 undergraduate and graduate law students to manually code the content of these 1492 cases.³⁹ This resulted in 2,186 manually-coded

³⁶ Historically, each Brazilian Tribunal has been responsible for its own IT systems, leading to significant discrepancies. TSE itself hosts a searchable public database (available at <http://www.tse.jus.br/jurisprudencia/decisooes/jurisprudencia>). but this database is/was not reliable. First, it was optimized to return fast queries and a maximum of 1000 results; this sacrifices accuracy, meaning that similar searches yield different results. Second, and most worrisome, the database requires judges to manually upload cases to TSE's system, which becomes particularly challenging when judges are overloaded with 24-hour deadlines. The differences in resources between jurisdictions could also indicate that results are biased as judges in more resourceful courts are more likely to upload cases than those in less resourceful courts. TSE judges themselves told us they do not trust their database.

³⁷ This effort initially resulted in a dataset of over 95,000 cases between December 2017 and March 2019, covering more than the entire 2018 electoral period. Second, CEPI-FGV developed a word search utilizing regular expressions of terms in three categories: (i) fake news, disinformation, and "known to be untrue" information; (ii) online content, digital content, and social media-related terms; (iii) date and time terms to restrict the focus to the 2018 elections. This resulted in over 850 useful expressions and a preliminary sample of 2,928 cases that match one wording at least in categories (i) and (ii). Finally, CEPI manually curated the cases to filter out any disputes that did not involve online "fake news" claims; that is, online content challenged for untruthfulness, fake news, disinformation, libel or similar. Because of the lack of strict definitions for what constitutes online disinformation, it incorporated a broad scope of language that minimally signifies that a content may be untrue. The final sample consists of **1,492 cases**, which should be close to the entire population of online misinformation cases decided by Brazilian courts during the 2018 Federal and State Elections

³⁸ The database was published in 2019 and is available at <https://direitosp.fgv.br/en/projetos-de-pesquisa/eleicoes-fake-news-tribunais-desinformacao-online-nas-eleicoes-brasileiras>. Rodrigo Karolczak, one of the authors of this article, was the main researcher in charge of developing the database, which relied on an unconditional grant from Facebook—meaning that Facebook funded the works that led to its creation but retained no overseeing powers over the final construction of the database.

³⁹ We developed a template for coding the cases. Then, we initially provided a sample of coded cases to teach students how to correctly label the cases. We then split students into groups to promote exchanges and add consistency within the groups. Finally, we held bi-weekly meetings with all students to discuss the coding and settle any potential discrepancies.

decisions⁴⁰. This is the universe of judicial interventions on disinformation during the 2018 Brazilian elections⁴¹, and it is remarkable in and of itself that: (i) a democratic poll fielding tens of thousands of candidates; (ii) where fake news was a central topic in the election; and (iii) where courts created a new right of action aimed explicitly at fighting disinformation, established broad standing rights and allowed any party to sue multiple times; led to *only* about 1500 cases and 2200 Court decisions over a roughly six-month period. All in all, Brazilian Courts removed from the internet around 1,602 URLs, out of 3,215 URLs properly identified in decisions themselves.

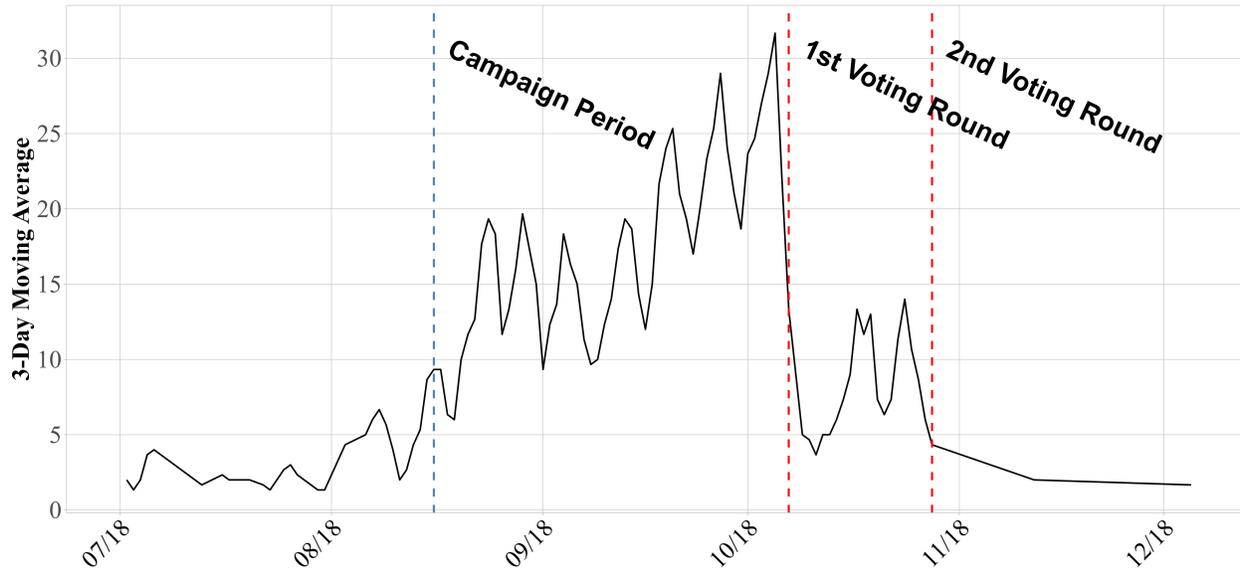
Below we present other key findings of this analysis.

General Trends: Figure 1 below plots the number of decisions on disinformation cases issued on a given day during the 2018 election period. As seen, litigation increased as the election's first and second voting rounds approached (October 7th and 28th, 2018, respectively). There was also a spike in cases in early August, which preceded the official start of the campaign period.

⁴⁰ Because of interim rulings, appeals, and others, one case may have multiple decisions.

⁴¹ Our 2018 manual codebook included information such as: (i) the involved parties (e.g. if plaintiffs are incumbent politicians and whether defendants are politicians, platforms, citizens or journalists); (ii) whether the decision resulted in content takedown, fines, rights of response, and/or bans on reposting similar content; (iii) The URLs challenged, the type of content (post, image, video, etc.), including whether the specific URL has been taken down or not; (iv) How plaintiffs characterized their claim (facts "known to be untrue," defamation, libel, slander, offense against honor) and how judges ruled on these claims; (v) Under what legal basis judges ordered the takedown of the content; (vi) Objective information on the judges (including gender and appointment process).

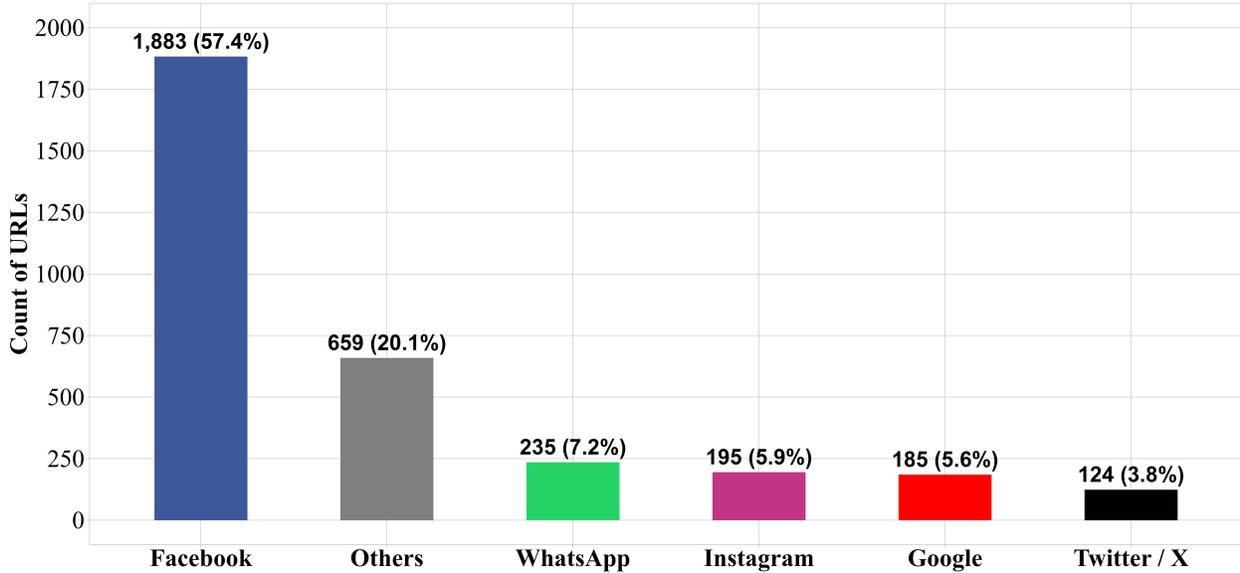
Figure 1: Timeline of filing dates for lawsuits on online disinformation during the 2018 elections



Graph displaying the 3-day moving average of filing dates for electoral lawsuits on online disinformation during the 2018 Brazilian elections. The majority of cases are introduced during the campaign period up to the 1st voting round, while the period before the 2nd voting round presents a smaller wave of cases.

Brazilian law required plaintiffs to identify the specific URLs that shared the alleged fake content as a requirement for filing a lawsuit. We can use this information to build a “market share” of alleged disinformation sharing according to the lawsuits (Figure 2). Facebook, WhatsApp, and Instagram accounted for 70.5% of all URL removal requests. Google/YouTube and Twitter/X only accounted for around 9.4% of URLs. Blogs, news platforms, and other websites accounted for approximately 20.1%.

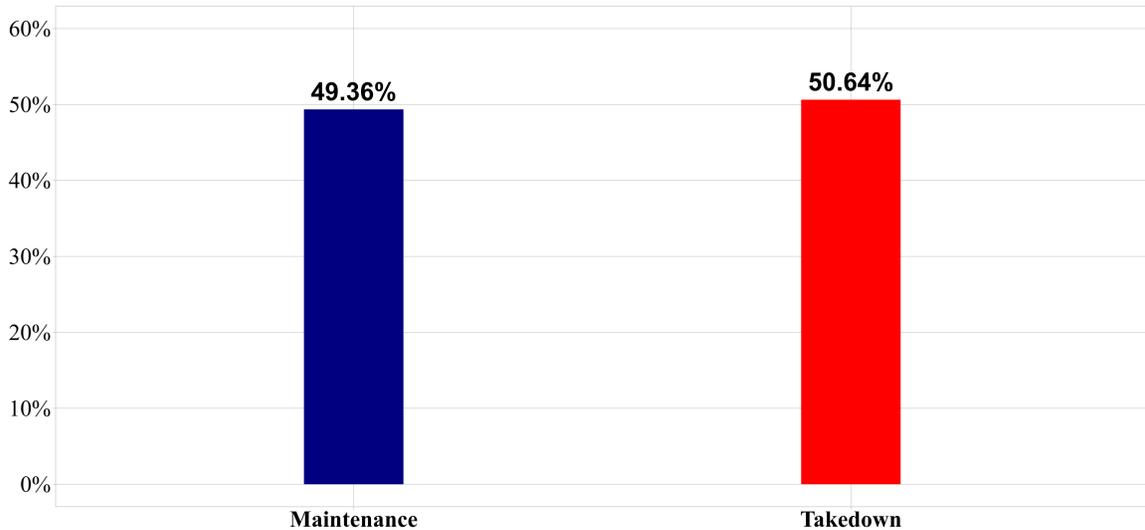
Figure 2: Count of sampled URLs in decisions by platform, 2018 elections



Graph depicting the count of sampled URLs alleged to contain disinformation during the 2018 elections by platform. Most of available URLs were in Facebook, followed by media in other websites such as small press and blogs.

In terms of outcomes, Brazilian judges ordered the takedown of content in 50.6% of cases,⁴² a significant percentage (Figure 3).

Figure 3: Takedown ratio for decisions during the 2018 elections

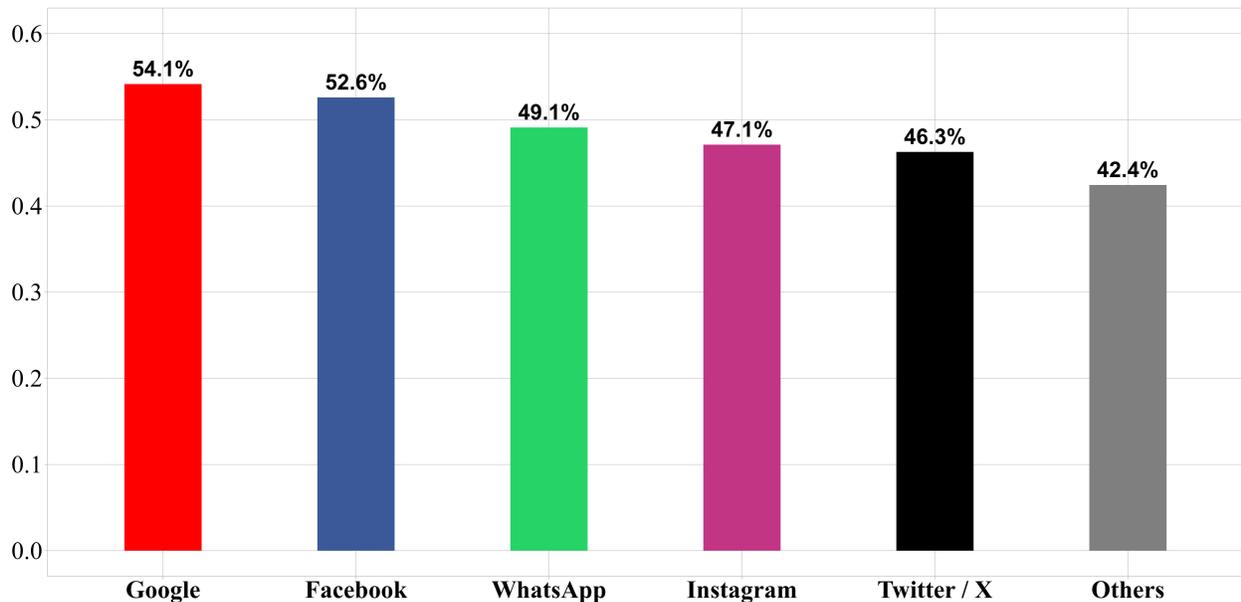


Graph showing the average maintenance rate and, conversely, the takedown rate of decisions involving requests for content removal during the 2018 elections. The N of decisions for this type of request is 2186, while 217 decisions did not involve a content removal request.

⁴² This figure considers “takedown” as the judge ordering the removal of any type of content in a lawsuit. It is not broken down per requested URLs as in Figure 2.

Potentially more worrisome, Brazilian Judges issued take-down orders in 54% of cases where the main defendants were ordinary citizens. Plaintiffs could request the takedown of an entire user profile (as opposed to a single post). Judges granted those profile takedown requests in 27.6% of cases. Plaintiffs could also sue other candidates running for office for spreading disinformation. In those involving two political candidates/parties, judges issued takedown orders in 43% of the lawsuits. Overall, most lawsuits targeted regular voters (876 out of 1,492 cases), rather than competing candidates (540 out of 1,492 cases). Figure 4 plots takedown ratios of URLs in decisions per platform, ranging from 42.4% (Others) to 54.1% (Google).

Figure 4: Takedown ratio of sampled URLs by platform, 2018 elections



Graph presenting the takedown rate of sampled URLs alleged to contain disinformation during the 2018 elections by platform.

Many have expressed concerns that combating disinformation will become more complex as users migrate toward encrypted peer-to-peer communication systems, such as WhatsApp, Telegram, or Signal⁴³. Interestingly, even back in 2018, WhatsApp already represented 8% of all challenged “URLs”. WhatsApp, however, does not have a URL system. Plaintiffs employed creative methods to circumvent this legal requirement, such as utilizing links to the WhatsApp Web version of the service. Additionally, the end-to-end encryption did not deter judges from issuing takedown orders in

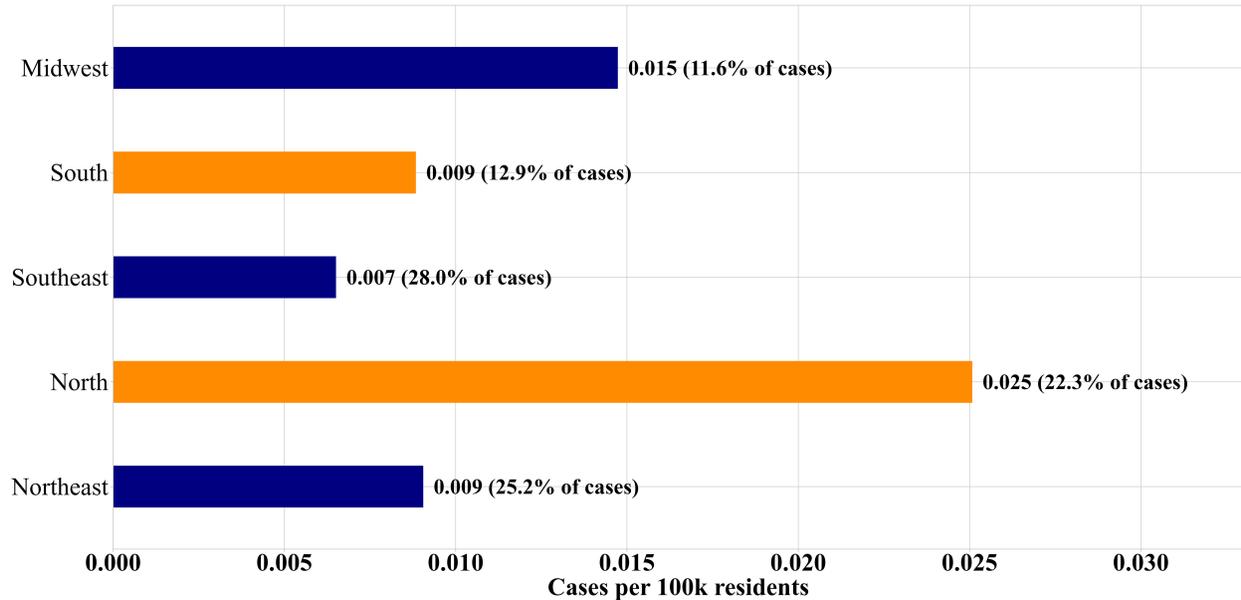
⁴³ See Tiago Ventura et al., *Misinformation Beyond Traditional Feeds: Evidence from a WhatsApp Deactivation Experiment in Brazil*, preprint J. POLIT. (2025). For a review of the literature and an interesting experiment involving the Brazilian elections.

approximately 41% of cases involving WhatsApp, although we do not have information on overall compliance levels with these decisions. This illustrates how plaintiffs can use creative means to get around procedural limitations, and how judges are willing to accept these innovations and expand powers to takedown content.

In terms of other punishments, despite their clear legal powers to impose fines for sharing disinformation and the overall 50% general takedown ratio, judges were reluctant to impose fines on parties in legal proceedings. In only 3% of cases, judges imposed a monetary fine on defendants for sharing disinformation; in 25% of cases, judges imposed daily fines if parties refused to comply with takedown orders. In other words, despite a vast number of available legal remedies, judges mostly resorted to notice-and-takedown as their punishment for the sharing of fake news.

Political Influence: In general, broad freedom of expression rights are justified on a fear of abuse by those in power: if a door is open for legislators or courts to control speech, elites and the government will suppress speech they dislike⁴⁴. Some of our evidence corroborates this fear: TREs in smaller, less economically developed states of Northern Region of Brazil (e.g., Amapá, Rondônia, Roraima, and Amazonas) received a disproportionately larger number of claims when compared to larger, more economically developed regions such as the Midwest and Southeast (e.g., São Paulo, Rio de Janeiro, and Minas Gerais—see Figure 5).

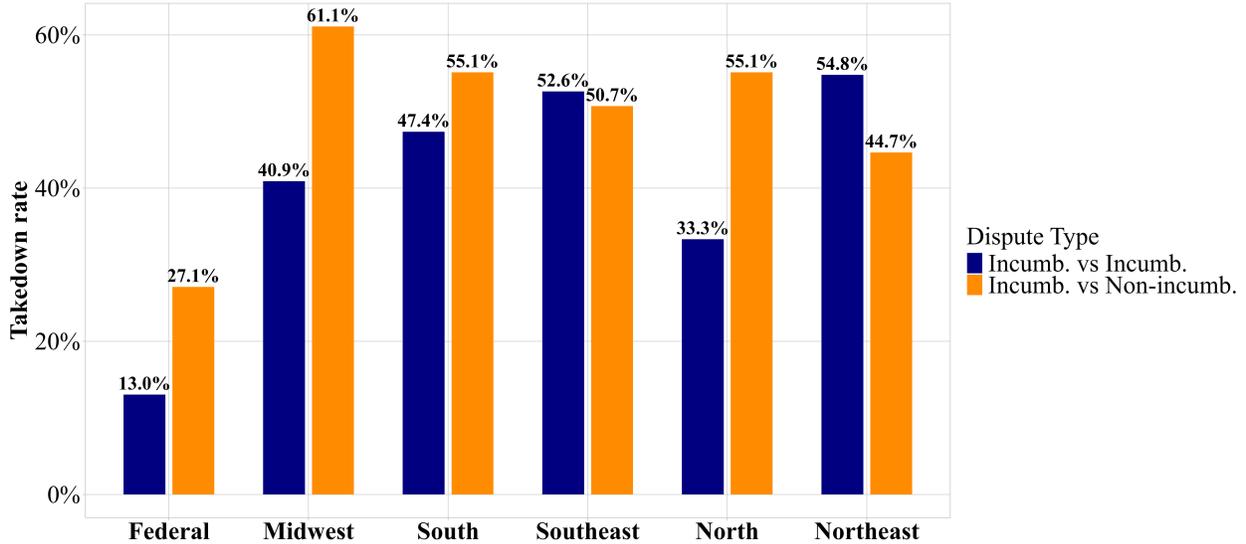
⁴⁴ Jeff Kosseff, *The Twenty-Six Words That Created the Internet*, in *THE TWENTY-SIX WORDS THAT CREATED THE INTERNET* (2019).

Figure 5: Cases of alleged online disinformation per 100k residents by region (ordered by GDP per capita)

Graph displaying the regional distribution of alleged online disinformation cases during the 2018 elections, scaled by population and ordered by average regional GDP per capita.

Nationwide, incumbents (considered those who occupied any public office in the year before the election) and non-incumbent politicians were equally likely to win takedown requests—both had win rates of 49%. Yet, these averages masked significant heterogeneity. As Figure 6 shows, incumbents were significantly more likely to win cases when they were suing non-incumbents than when they were suing other politicians already in office in many Brazilian States. This evidence can indicate some form of capture by politicians already in power. Incumbents also used the system much more—they were the plaintiffs in approximately 1,180 cases, versus 395 for non-incumbents.

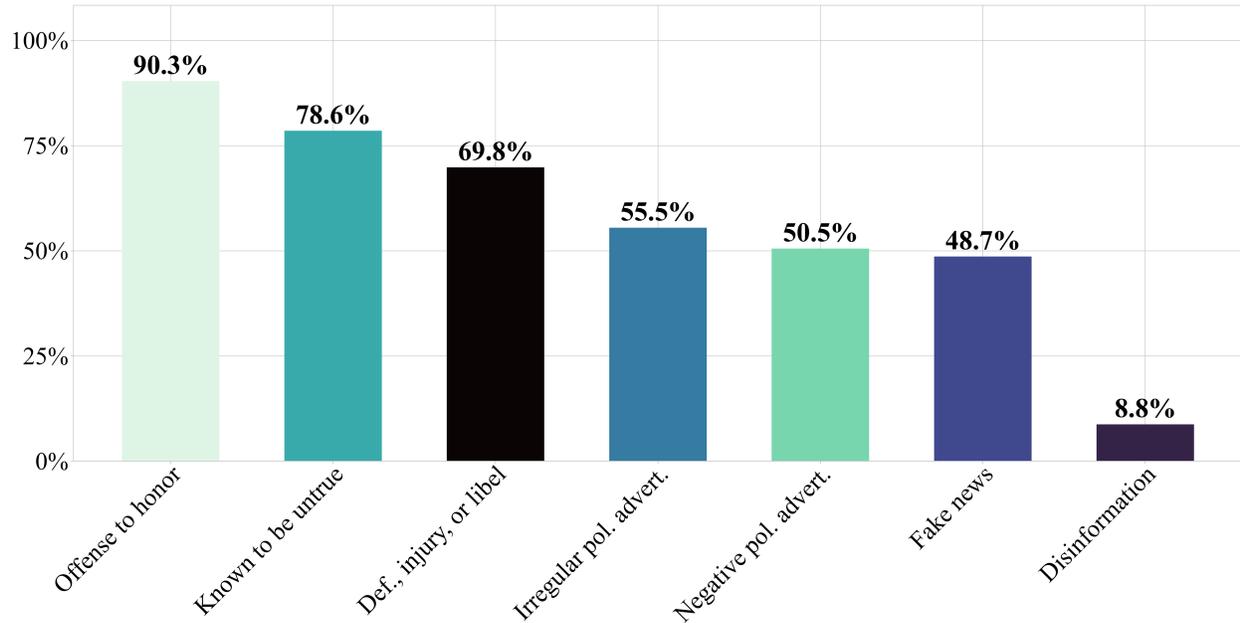
Figure 6: Takedown rates by plaintiff and defendant incumbency by region (ordered by GDP per capita)



Graph showing the regional distribution of legal disputes over online disinformation between incumbents and non-incumbents during the 2018 elections. Incumbency is defined as holding public office up to one year prior to the electoral year. 1172 decisions involve an incumbent plaintiff versus a non-incumbent defendant, while 253 decisions involve incumbents as plaintiffs and defendants.

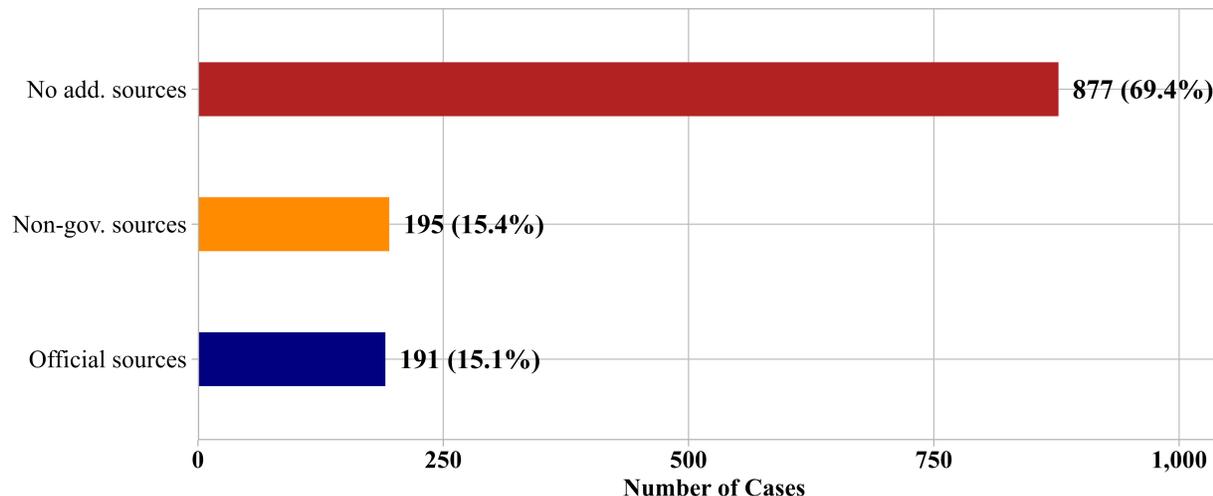
Legal Standards: Some jurisdictions, such as Germany, have opted to permit the removal of content only when the underlying material also constitutes a crime. Our data indicates that this is a poor filter. In 90% of Brazilian cases, plaintiffs also alleged offense to honor, and 70% of claims combined accusations of disinformation with a punishable crime, such as defamation or libel (which may have some overlap with “offense to honor” under Brazilian Law). Interestingly, the colloquial term “fake news” had a significantly larger share than a technical term like “disinformation,” which speaks to the novelty of the problem in 2018.

Figure 7: Terms employed by judges when adjudicating disinformation cases, 2018 elections



Graph depicting the language and terms used by cases on online disinformation during the 2018 elections.

The 2018 version of the Brazilian system, however, did not appear to have developed an alternative, coherent legal standard for determining what types of content qualify as disinformation or fake news. In 53% of rulings, judges did not quote a *single piece of case law* when justifying their decisions, a proxy for the quality of argumentation in these decisions. In addition, in 70% of cases, judges did not cite or add any additional information or evidence to their rulings while discussing whether the content shared “facts *known* to be untrue”, the legal standard that should rule decisions on disinformation cases (figure 8). Judges also reversed only 15% of their initial interim rulings, which were typically made within 24 hours of the complaint's filing, and often without defendants having the opportunity to present their case to the Court. Only 16% of cases were appealed by the losing party, and higher Courts rejected 72% of these appeals.

Figure 8: Citations to sources when adjudicating disinformation cases, 2018 elections

Graph illustrating how the concept of ‘facts known to be untrue’ was applied in decisions on online disinformation during the 2018 elections. A total of 1,382 decisions employed this concept, though its application varied significantly. We identify three distinct standards of truth: (i) those grounded in official government documents, (ii) those relying on non-governmental sources such as the press, and (iii) those lacking any additional sources.

On the other hand, takedown rates were significantly smaller when judges conducted research and cited case law in their decisions (a 40.9% takedown rate) compared to when they did not (a 59.3% rate).

In Appendix I, we use different Logit models to measure the effect of a judge citing key provisions of Brazilian electoral advertising law to understand the strength of these correlations. Results indicate that citations to the law increased the odds of content removal by approximately 2 times, with an estimated probability of removal of 67.8%. The models suggest that whenever a judge cited any other case law, the odds of content removal fell by approximately 0.44 times, with an estimated probability of removal of 31.3%. This may signal either that these are more complex cases that require additional research, that the simple fact of researching the legal standards makes judges more reluctant to take down content, that judges believe they need bolder justifications to keep content online, or a combination of these situations.⁴⁵ It is noteworthy that these are professional, well-trained, full-time federal and state judges who are accustomed to a system of precedents and well-reasoned decisions that are almost always appealed to higher courts. Issuing a court order with only limited reasoning is a departure from their usual practice; however, it represented the majority of online fake news decisions nonetheless.

⁴⁵ Our data does not allow us to disentangle these explanations or to assess the underlying veracity of the shared content.

b. Disinformation Adjudication in the 2022 Brazilian Elections

Fake news was again at the center of the 2022 Brazilian Presidential dispute,⁴⁶ now between President Bolsonaro running for reelection against former president Lula. Like President Trump in 2020 and 2024, Bolsonaro started attacking the integrity of the Brazilian voting process before the vote itself took place, claiming that the system would be manipulated to ensure his defeat. Despite being President, Bolsonaro did not provide any evidence to support his claims against the system—on the contrary, the audits he ordered reaffirmed the integrity of the elections.⁴⁷ Lula ultimately won the tightest-ever presidential vote in Brazil by a 51%-49% margin, putting the left-wing Workers' Party back in power.

Our approach to tracking disinformation litigation during this election differed from the ex-post comprehensive survey of the 2018 case law. For 2022, FGV-CEPI and FGV-COMPPIT established a joint project—the FGV Law School Sao Paulo Disinformation Observatory for the 2022 elections—where 16 researchers followed the week-to-week evolution of court cases and their decisions throughout the election calendar.⁴⁸ The project ran a web scraper

⁴⁶ Mac Margolis & Robert Muggah, *Opinion | Brazil's Fake-News Problem Won't Be Solved before Sunday's Vote*, THE WASHINGTON POST, Oct. 27, 2022, <https://www.washingtonpost.com/opinions/2022/10/27/brazil-election-bolsonaro-lula-fake-news/>.

⁴⁷ President Bolsonaro ordered the army to audit the safety of the electronic voting machines used in all Brazilian elections, and the army's audit as well as numerous other audits found that the machines were safe and reliable—the underlying code is open source and the machines, for example, never connect to the internet, so the tens of thousands of machines used in a given poll would need to be hacked one by one, despite being heavily guarded before, during and after the poll. That did not stop Bolsonaro from attacking the system nonetheless.

⁴⁸ The process of selecting and categorizing the decisions involved several criteria, focusing on key elements of different cases. These included the type of judicial decision (i.e., interim or final), the parties involved, the facts and legal grounds set forth by claimants, and the merits of the case as established by the decision. Each decision was analyzed in terms of its relevance, considering the jurisdiction (the higher court – TSE – or state-level courts – TRE) and the type of measure applied (removal, compensation, right of reply, among others). Additionally, researchers evaluated the legal reasoning supporting the decisions, including citations to relevant legal provisions, similarly to what was done for the 2018 elections.

Caio Mario and Rodrigo Karolczak led this project, together with Professor Alexandre Pacheco from FGV-CEPI. The results of the observatory were posted online during the election itself, with weekly reports about the key cases under adjudications. The researchers also participated in webinars aimed at discussing the project and contributed sporadically with press coverage by answering interviews from journalists before, during, as well as after the elections. FGV Direito SP made an agreement with Folha de São Paulo, a large newspaper based in São Paulo, to provide real-time information and analysis about how electoral courts were dealing with disinformation. See a description of how this coverage was organized here:

in search of new cases and manually searched for upcoming cases to identify key decisions that shaped how courts tackled electoral disinformation while the election itself was ongoing.⁴⁹

Brazilian electoral courts considered themselves to be better prepared to fight disinformation in 2022.⁵⁰ They implemented the same system as in 2018, with two important exceptions:

- i. Courts expanded their power to adjudicate disputes, now relying on two alternative legal standards: (i) the older “*facts known to be untrue*”; and (ii) a new standard that prohibited the sharing of “*seriously decontextualized*” information⁵¹.
- ii. In between the first and second round of the 2022 elections, the Courts published⁵² a new regulation: (i) establishing additional sanctions to the dissemination of *facts known to be untrue* and/or *gravely decontextualized information* that targeted specifically “*the integrity of the electoral process*”; (ii) granting the president of the Superior Electoral Tribunal the power to order the removal of content “identical to other previously illegal content”, therefore weakening the URL based-system and enabling the en masse targeting of similar messages; (iii) empowering courts to suspend profiles, accounts and channels involved in the “*production of systematic disinformation, characterized by the repeated publication of false and decontextualized information about the electoral process*”; and (iv) prohibiting any paid dissemination of electoral propaganda on the internet.

FGV Direito Sao Paulo, *FGV Direito SP e Folha Acompanharão Trânsito de Decisões Sobre Desinformações Nas Eleições de 2022*, (Aug. 2022), <https://direitosp.fgv.br/noticias/fgv-direito-sp-folha-acompanharao-transito-decisoos-sobre-desinformacoes-nas-eleicoes-2022>.

⁴⁹ Some decisions were not caught on the weekly scrapping, either because they had not been published yet at the Official Gazette or because of technical difficulties with the scraper and filters used. In any event, the researchers also followed the press releases of the superior courts and manually gathered relevant decisions not caught in the scraping process.

⁵⁰ Cristina Tardáguila, *One Year Later: Brazil Has Never Been So Prepared to Fight Fake News*, AMERICAS QUARTERLY (Sept. 12, 2022), <https://www.americasquarterly.org/article/one-year-later-brazil-has-never-been-so-prepared-to-fight-fake-news/>.

⁵¹ TSE. Resolution No. 23.610, of December 18, 2019, updated by Resolution No. 23.671, of December 14, 2021, which inaugurated the concept of “seriously decontextualized”, both available at: <https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019> and <https://www.tse.jus.br/legislacao/compilada/res/2021/resolucao-no-23-671-de-14-de-dezembro-de-2021>

⁵² Resolution No. 23.714, of October 20, 2022. Available at: <https://www.tse.jus.br/legislacao/compilada/res/2022/resolucao-no-23-714-de-20-de-outubro-de-2022>

Overall, the Disinformation Observatory identified and analyzed 427 decisions connected to the 2022 electoral voting process. We are confident that our database covers the vast majority of key decisions, even if we cannot provide comprehensive evidence as we did for 2018. Overall, 331 decisions addressed content and profile removal requests: 155 decisions (or 47% of this sample) denied the removal request, finding the claims incompatible with either of the two legal standards; while 176 decisions (53% of the sample) ordered takedowns⁵³.

Judges denied takedown orders *without citing any additional information source* in 54% of the cases in our sample, with the remaining 46% of decisions citing news pieces from trusted sources, official documents, and other versions of the content under discussion. Out of the 176 decisions ordering a takedown, 63% mentioned such additional sources of information. This data represents an increase in citations to additional sources of information between 2018-2022, though we cannot rule out that this was caused by sample selection, as the Observatory was naturally looking for higher-profile cases that are more likely to undergo in-depth analysis.

Even in 2022, however, Judges continued to struggle to define what standard should guide decisions on what facts are “*known to be untrue*” or are “*gravely decontextualized*”. In particular, a comprehensive analysis of these hundreds of key cases indicates the following developments:

- Courts sometimes relied on fact-checking agencies to justify their decisions, but this was not a recurrent pattern—in many rulings, judges merely invoked prior knowledge of the subject matter when adjudicating the dispute. This was especially true in the large number of rulings that did not cite any additional informational sources;
- Courts took steps to protect satirical or humorous content, even when they decontextualized information. However, courts did not produce a consistent criterion to define satirical content that merited such a flexible standard, leaving room for significant discretion between judges and cases;
- Some decisions protected traditional media from intervention, especially when judges found evidence that the news outlets investigated the underlying facts before publishing the challenged story. However, courts also sometimes held professional journalists to higher standards, requiring them to report accurately and ordered the removal of content that contained “technical” mistakes such as a wrong reference to the legal

⁵³ For more specific data, see “Os Números da Desinformação na Justiça Eleitoral”, in “Disinformation Observatory in 2022 Elections”, available at <https://medium.com/observat%C3%B3rio-da-desinforma%C3%A7%C3%A3o-nas-elei%C3%A7%C3%B5es-2022/os-n%C3%BAmeros-da-desinforma%C3%A7%C3%A3o-na-justi%C3%A7a-eleitoral-495d597ee0e6>

status of a criminal investigation (e.g. an inquiry versus a criminal charge);

- Courts continued to struggle to deal with peer-to-peer messaging platforms such as WhatsApp and Telegram. Sometimes, judges ruled that the private nature of the underlying communication removed such platforms from the scope of electoral regulation. In other cases, they ordered the takedown of videos circulated in WhatsApp groups and the posting of the judicial decision in the same groups where the video was posted, deeming such groups as quasi-public in nature;
- Some decisions identified the creation of a broader ecosystem to spread disinformation across platforms. When this was the case, courts decided not only to take down specific content but also to order social media platforms to suspend the monetization of specific channels and profiles, limiting payments to content creators during the election period; and, finally
- Electoral courts were particularly concerned with the use of manipulated videos or images to spread information, regularly requiring the take down of information that involved AI (e.g., deep-fakes) or other forms of manipulation;

Overall, the decisions taken during the 2022 elections did not substantially change the pattern identified in the more comprehensive analysis of 2018. The major changes were the inclusion of a new legal standard (*gravely decontextualized information*) and the expansion of the courts' powers to take down similar content already considered illegal in other decisions and to fight more systematic disinformation (especially disinformation regarding *the integrity of the electoral process*). The everyday judicial practice, however, remained broadly similar.

c. The 2022-2025 Aftermath: an increasingly powerful “content-moderation” Court

As described above, Brazilian Courts initially designed this evolving system to tackle the spread of disinformation during election periods, relying on their powers to police electoral advertising. This goal shaped many institutional design choices, such as the expedited procedural requirements—with decisions sometimes issued in under 24 hours—and the overall limitation that all take-down orders are automatically terminated once a given election concludes.

However, over time, the Brazilian Superior Electoral Court and the Brazilian Supreme Court expanded their powers to police the sharing of disinformation well beyond the electoral period. For example, in March 2019, the Supreme Court initiated an official investigation into what it perceived as a

growing disinformation campaign targeting the Supreme Court itself, its members, and their family members.⁵⁴ This investigation, which became known as the “Fake News Inquiry”, has been ongoing for more than six years as of the writing of this article. In July 2021, the Supreme Court opened the “Digital Militias Inquiry” to investigate the alleged existence of a criminal organization using digital platforms to attack Brazilian democracy and the rule of law, including through the dissemination of disinformation—another investigation that has been ongoing for more than four years.⁵⁵ The broad scope of these and other cases, which focus on attacks against the Court, Brazilian democracy, or the rule of law, means that the Court has effectively granted itself the power to police a large swath of information dissemination in online platforms within and outside of election periods.

These investigations are also based on ad hoc procedural rules. More specifically, because they were opened by the Court rather than the Attorney General, they are being directly supervised by the Justices themselves, who have the power to direct police investigations and issue rulings—a breach of the traditional separation between prosecutorial and adjudicating functions. Both these and other investigations have been centralized in Justice Alexandre de Moraes, a former prosecutor, who, as Justice rapporteur, has broad powers to order dawn raids or require the takedown of content or social media profiles. Justice Moraes’ decisions can be appealed to a panel of five Supreme Court Justices (including him), and then to the full court of 11 Justices, but there are many procedural ways to delay this review. Both the panel and the Court have largely upheld Justice Moraes’ decisions, regardless of their controversy. Over the years, Justice Moraes issued decisions requiring five social networks (i.e. Discord, Meta, Rumble, Telegram and Twitter) to suspend the profiles and channels of the influencer Bruno Monteiro Aiub (also known as Monark), for disseminating hate speech and attacking the institutional and democratic order,⁵⁶ and requiring OnlyFans to take down the account of a self-exiled Brazilian blogger based in the US—Allan do Santos—that recurrently attacks the integrity of Brazilian elections and the Supreme Court.⁵⁷ These were all controversial rulings that placed Justice Moraes and the Supreme Court squarely at the center of Brazilian political debates. Many of these decisions were confidential

⁵⁴ BRAZIL. Supreme Court, INQ 4.781, Reporting Minister Alexandre de Moraes, decision of May 26, 2020. Available at: <https://www.stf.jus.br/arquivo/cms/noticiaNoticiaStf/anexo/mandado27maio.pdf>.

⁵⁵ BRAZIL. Supreme Court, INQ 4.874, Reporting Minister Alexandre de Moraes, decision of July 16, 2021. Available at: <https://portal.stf.jus.br/processos/downloadPeca.asp?id=15347088212&ext=.pdf>.

⁵⁶ See <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=508935&ori=1>

⁵⁷ See Maria Magnabosco, *Moraes manda bloquear OnlyFans de blogueiro bolsonarista foragido da Justiça*, ESTADÃO (July 10, 2025), <https://www.estadao.com.br/politica/alexandre-de-moraes-manda-bloquear-onlyfans-blogueiro-bolsonarista-foragido-justica-nprp/>.

takedown orders or other orders issued directly to the social media platforms that hosted such accounts. Still, as ad hoc as these procedures are, these decisions are subject to appeal and are legally valid and binding orders of the Brazilian Supreme Court.

The dispute between the Brazilian Courts, Justice Moraes, and the digital platform Twitter/X took place against the backdrop of an increasingly powerful and politically polarizing Court. In a way, it works as a case study of the evolution of the Brazilian Court-led content moderation and disinformation adjudication system.

On January 8, 2023, one week after President Lula took power after winning the 2022 elections, violent riots similar to the January 6 attacks on the US Capitol struck Brasilia, the Brazilian national capital. They culminated in the invasion and partial destruction of the Brazilian Congress, Presidential Palace, and Supreme Court, and are nowadays referred to as the “January 8 Riots.”⁵⁸ These riots triggered a large-scale investigation by the Brazilian Federal Police on threats to the national democracy that uncovered an alleged attempted Coup d’Etat against the newly elected Brazilian government. In early 2025, Brazilian prosecutors filed criminal lawsuits against 34 individuals before the Brazilian Supreme Court, with former President Jair Bolsonaro listed as one of the leaders of the attempted Coup.⁵⁹ The trial is ongoing as of the drafting of this article. Still, senior military officials agreed to a plea deal where they are confirming the attempted Coup in exchange for lower sentences.

In April 2024, while the Federal investigations into the January 8 Riots were ongoing, the Brazilian Supreme Court opened a parallel procedure (PET 12404) to scrutinize alleged threats and a disinformation campaign against the Federal Police officers leading the investigations⁶⁰. Also in April 2024, X allowed some user accounts that the Supreme Court had previously ordered to be taken down as part of the Digital Militias Inquiry to transmit live content where these accounts criticized the new PET 12404 parallel procedure as an attack on freedom of speech by opposition members. Remember that the Digital Militias Inquiry was initially opened by the Court in 2021 to investigate whether far-right wing bloggers were using social media networks to spread disinformation in a way that threatened Brazilian democracy by challenging the reliability of the poll. It is not surprising that the same bloggers were now attacking the investigation of the January 8 Riots as a politically motivated

⁵⁸ Jack Nicas & André Spigariol, *Bolsonaro Supporters Lay Siege to Brazil’s Capital*, THE NEW YORK TIMES, Jan. 9, 2023, <https://www.nytimes.com/2023/01/08/world/americas/brazil-election-protests-bolsonaro.html>.

⁵⁹ Jack Nicas, *Brazil Charges Bolsonaro With Plotting a Coup After 2022 Election Loss*, THE NEW YORK TIMES, Feb. 19, 2025, <https://www.nytimes.com/2025/02/18/world/americas/brazil-bolsonaro-coup-charges.html>.

⁶⁰ BRAZIL. Superior Tribunal Federal, PET 12404, Reporting Justice Alexandre de Moraes. Available at: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=6888934>.

witch-hunt—the same dynamics played out in the US, as the January 6 invasion of the US Capitol turned from a “violent mob” to a “day of love.”⁶¹ Importantly, at least one of the persons impacted by these orders and investigations is a Brazilian blogger (Allan dos Santos) who went into self-exile in the US in 2020. Mr. dos Santos is a Brazilian citizen who lives in Florida and regularly posts content in Portuguese, targeting Brazilian citizens, from the United States. Justice Moraes ordered the arrest of Mr. dos Santos back in 2021 as part of the “Fake News” inquiry. Brazilian authorities asked the US to extradite Mr. dos Santos, but the US government has not complied with this request⁶².

The Brazilian Supreme Court, then, always under the sole reporting of Justice Alexandre de Moraes, opened a separate proceeding (INQ 4957) to investigate whether X, now under the control of Elon Musk, had openly defied a lawful Brazilian Court order by allowing these previously blocked accounts to share content⁶³. X explained to the Court that a technical glitch led to the temporary, involuntary non-compliance with the previous Court order.

Starting in August of 2024, the clash between the two parties escalated quickly:

- On August 7, Justice Moraes issued confidential takedown orders that required X to block new profiles created by some of the users previously barred from X as part of these disinformation investigations (including Mr. dos Santos). These orders subjected X to a substantial daily fine in the event of non-compliance with the Court order.⁶⁴ X refused to implement those orders, challenging their legality and keeping the accounts online;
- On August 17, 2024, X announced the shutdown of its Brazilian office—a decision that left the company without any legal representative in the country, in an attempt to avoid enforcement actions against the company and its representative⁶⁵. It is noteworthy

⁶¹ Ryan Reilly, *Trump Once Called Jan. 6 a “heinous Attack.” Now He Calls It a “Day of Love.”* NBC NEWS (Oct. 17, 2024), <https://www.nbcnews.com/politics/donald-trump/trump-called-jan-6-heinous-attack-now-calls-day-love-rcna175942>.

⁶² For a summary, see *Allan Dos Santos*, WIKIPEDIA (2025),

https://en.wikipedia.org/w/index.php?title=Allan_dos_Santos&oldid=1299860595.

⁶³ BRAZIL. Superior Tribunal Federal, INQ 4957 - DF, Reporting Justice Alexandre de Moraes. Available at: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=6893258>.

⁶⁴ BRAZIL. Superior Tribunal Federal, PET 12404, Reporting Justice Alexandre de Moraes, decision of August 30, 2024, pg. 2-3. Available at: <https://noticias-stf-wp-prd.s3.sa-east-1.amazonaws.com/wp-content/uploads/wpallimport/uploads/2024/08/30171714/PET-12404-Assinada.pdf>.

⁶⁵ Jack Nicas, *Elon Musk Closes X Office in Brazil Over Fight With Judge*, THE NEW YORK TIMES, Aug. 18, 2024, <https://www.nytimes.com/2024/08/18/world/americas/elon-musk-x->

that Brazilian Courts had previously ordered short-term arrests of high-level executives from both Google (in 2012) and Meta (in 2016) as a means to ensure that large digital platforms comply with Brazilian Court orders.

- On August 18, Justice Moraes ordered banks and the Brazilian government to block all accounts and assets of X in Brazil, and ordered X to appoint a new representative in the country within 24 hours, threatening the suspension of all its activities in the country in case of non-compliance. X again refused to comply with the Supreme Court order. Justice Moraes also ordered the freezing of all accounts and assets of Starlink in Brazil, despite Elon Musk's satellite internet company being part of SpaceX, a separate legal entity from X⁶⁶.
- On August 30, Justice Moraes ordered the Brazilian National Telecommunications Regulator (ANATEL) and all Brazilian internet service providers to suspend access to X's IP addresses in all Brazilian territory until the company complied with the Court orders: appointing a new legal representative, suspending the required X accounts, and paying all outstanding fines.⁶⁷ Starlink, Elon Musk's satellite internet provision company, initially refused to comply with the order to deny access to X, effectively bypassing the Brazilian Supreme Court requirement by using its space-based infrastructure to beam internet to Brazil. Starlink eventually complied with the order after ANATEL threatened to seize Starlink's ground assets in the country.⁶⁸ The panel responsible for overseeing appeals against Justice Moraes' orders quickly and unanimously confirmed the suspension of X's IP addresses.
- On August 31, a new account hosted by X called @alexandrefiles started releasing Justice Moraes' confidential orders. X itself apparently controlled the account, and its name connected Justice Moraes' decisions with the so-called Twitter Files released by Mr. Musk to prove how the Biden administration was using

brazil.html. Elon Musk commented the decision to shut down the office on X (<https://x.com/GlobalAffairs/status/1824819053061669244>).

⁶⁶ Associated Press, *Elon Musk's Starlink Backtracks to Comply with Brazil's Ban on X*, THE GUARDIAN, Sept. 4, 2024, <https://www.theguardian.com/technology/article/2024/sep/04/elon-musk-x-starlink-brazil>.

⁶⁷ BRAZIL. Superior Tribunal Federal, PET 12404, Reporting Justice Alexandre de Moraes, decision of August 30, 2024. Available at: <https://noticias-stf-wp-prd.s3.sa-east-1.amazonaws.com/wp-content/uploads/wpallimport/uploads/2024/08/30171714/PET-12404-Assinada.pdf>.

⁶⁸ Press, *supra* note 66.

governmental powers to silence conservative voices in the US.⁶⁹ The @alexandrefiles account accused Justice Moraes of recurrently ordering the secret censoring of prominent right-wing politicians and journalists, and everyday Brazilians, including many who resided in the United States (such as Mr. dos Santos).

- In late September, X retreated and complied with the orders to take down the accounts and appoint a new legal representative, but refused to pay the fines for contempt of Court that ultimately amounted to BRL 28.6 million (equivalent to USD 5.3 million at the time). On September 27, the Supreme Court recognized compliance with the court orders but required the payment of the fines to lift the blockage. On October 4, X paid all fines, and on October 8, the Brazilian Supreme Court reinstated access to the platform. Overall, all Brazilian users' access to X in Brazil was effectively suspended between August 30 and October 8.

The timeline below summarizes the main events of this dispute, with a special focus on the decisions that led to the blocking of X (indicated by the blue points in the timeline).

Figure 9 – Timeline of Procedures Involving Post-Election Misinformation (2024)



Elaborated by the authors

This, however, is not the end of the story. As mentioned in the introduction, in early 2025, Trump Media and the platform Rumble sued Justice Moraes in U.S. Federal Courts, accusing him of censoring the lawful speech of American citizens by issuing extraterritorial orders that impact US companies

⁶⁹ Aimee Picchi, *Twitter Files: What They Are and Why They Matter* - CBS News, CBS NEWS (Dec. 14, 2022), <https://www.cbsnews.com/news/twitter-files-matt-taibbi-bari-weiss-michael-shellenberger-elon-musk/>.

and residents⁷⁰. This litigation is ongoing. In July of 2025, President Trump ordered the imposition of 50% tariffs on US imports from Brazil—explicitly singling out Brazil’s content moderation regime and Justice Moraes as violating US freedom of expression and broader human rights laws⁷¹. This dispute is also ongoing, as President Lula has openly stated that the Brazilian judiciary is an independent body that is not accountable to him⁷².

Overall, this episode illustrates how the Brazilian Supreme Court has significantly expanded its powers to address all forms of content it deems “disinformation”. Investigations initiated during the elections were kept open after the electoral period. Other investigations, such as the one involving the online disinformation campaign against the federal police, were opened after the elections and have remained open ever since. In addition, because these procedures were centralized in the Supreme Court, the room for appeals is inherently limited, and procedural twists can further restrict it.⁷³ A single reporting Justice, Alexandre de Moraes, effectively concentrated the powers to issue injunctions and order any enforcement measures addressing issues broadly related to information that may impact the Brazilian democracy or the reputation of the Supreme Court.

It also showcases the complexity of governing online speech and the power of leading digital platforms. In many ways, Brazilian orders targeting U.S. residents are extraterritorial, even if not targeting them would create large loopholes in content moderation. In addition, Elon Musk and X unilaterally refused to comply with multiple official orders issued by the Brazilian Supreme Court, despite Brazil being the world’s 10th-largest economy and 7th-largest country in terms of population. This non-compliance continued for months despite the quick escalation of the dispute.

⁷⁰ Goodman, *supra* note 5.

⁷¹ President Donald J. Trump, *supra* note 6.

⁷² Jack Nicas & Victor Moriyama, *Brazil’s President Lula Voices Frustration With Trump Amid Tariff Battle*, THE NEW YORK TIMES, July 30, 2025, <https://www.nytimes.com/2025/07/30/world/americas/brazil-president-lula-trump-tariffs.html>.

⁷³ For example, X filed a writ against a decision of Justice Alexandre de Moraes, but the Supreme Court dismissed the writ based on procedural arguments, justifying that no writ was available against Supreme Court orders. See Felipe Recondo, *Opção da Starlink por mandado de segurança foi estratégica e marca posição contrária à decisão de Moraes*, (Sept. 4, 2024), <https://www.jota.info/stf/do-supremo/opcao-da-starlink-por-mandado-de-seguranca-foi-estrategica-e-marca-posicao-contraria-a-decisao-de-moraes>.

IV. Assessing the Brazilian Case

a. Five theoretical criteria to assess the Brazilian case

Designing an effective, transparent, and fair content-moderation system is a highly complex task, full of challenges and potential perils to freedom of speech. Thus, to assess the Brazilian system, it is helpful to have a clear framework in mind, mapping key points of concern. In a seminal article, Jack Balkin lists four main challenges that any governance system for online speech must overcome to be considered fair and effective:⁷⁴

- (i) developing effective standards that are protective of freedom of expression;
- (ii) avoiding a global jurisdiction (or some form of Brussels' effect⁷⁵ where one country effectively imposes standards on all others);
- (iii) preventing an obscure privatized bureaucracy from serving as prosecutor, judge, and executioner without accountability; and
- (iv) preventing nation-states from co-opting private infrastructures for surveillance, data collection, and analysis to increase their control over civil society⁷⁶.

To those, we must add the importance of responsiveness in remedy design and implementation:

- (v) Elections are dynamic, so tackling disinformation during election periods requires responsiveness and procedural flexibility. Similarly, large digital platforms have a long history of successfully evading regulatory regimes, so the system must be designed in a way that helps prevent such evasion.⁷⁷

These five criteria enable us to evaluate regulatory systems designed to tackle the negative impacts of disinformation on public spheres, during and outside of election periods. In particular, it is essential to emphasize the role of proper due process rights and procedures in ensuring that decisions around content

⁷⁴ Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUMBIA LAW REV. 2011 (2018). at 22.

⁷⁵ On the Brussel's effect, see Anu Bradford, *The Brussels Effect*, 107 NW UL REV 1 (2012)., on how global competition can lead to harmonization or capture, see Lancieri, Edelson, and Bechtold, *supra* note 9.

⁷⁶ On these final two, see also Evelyn Douek, *The Tug between Private and Public Power Online*, 61 DUQ REV 209 (2023).

⁷⁷ On the effectiveness of remedy design, see, Lancieri and Pereira Neto, *supra* note 23.

moderation are fair and appropriate.⁷⁸ However, these problems also require effective remedies—digital law is full of examples of paper tiger regulations that look strong on paper but fail to deliver any meaningful change for users and societies.⁷⁹

b. Applying these criteria to the Brazilian court-based content moderation system

From the perspective of a democratic theory of freedom of speech, Courts may be the right locus (or the least bad one) to tackle disinformation and preserve the public sphere. If the intention of the Law protecting freedom of speech is “*to broaden the terms of public discussion as a way of enabling common citizens to become aware of issues before them and of the arguments on all sides and thus to pursue their ends fully and freely,*”⁸⁰ judicial intervention against disinformation can preserve the overall fairness of the process, avoiding distortions in the information environment.

The Brazilian model—at least as designed initially—has many theoretical strengths:

- i. It is an open (multiple parties may file complaints) and appealable system, where independent/non-partisan, career judges and prosecutors decide what content should be taken down. This can allow for the development of clear standards on what is acceptable online speech through the evolution of case law;
- ii. The inherent publicity of court decisions allows some form of democratic accountability of the judicial standards—both civil society and Congress may criticize the results and even pass laws to correct interpretations. Appeals provide an internal avenue to correct mistakes;
- iii. Broad standing rights and Court adjudication of disputes prevent social media platforms from becoming the unique arbiters of online speech. Theoretically, the Judiciary can develop legal standards that are subsequently implemented by online platforms at scale;

⁷⁸ See Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV REV 526 (2022); Daphne Keller, *The Rise of the Compliant Speech Platform*, LAWFARE (2024), <https://www.lawfaremedia.org/article/the-rise-of-the-compliant-speech-platform>.

⁷⁹ See, for example, Filippo Lancieri, *Narrowing Data Protection’s Enforcement Gap*, 74 MAINE LAW REV. 15 (2022). on how digital platforms can capture regulators and hollow out regulatory interventions.

⁸⁰ OWEN FISS, *THE IRONY OF FREE SPEECH* (2009).

- iv. The reliance on the Judiciary (as opposed to regulatory agencies) diminishes the risk that the party in power abuses its authority to illegally prosecute political opponents;
- v. Streamlined procedural rules—most decisions are made between 24 and 72 hours after the complaint is filed—empower the Judiciary to act quickly, potentially thwarting electoral disinformation campaigns at their incipiency;
- vi. Takedown orders are restricted to Brazil and are no longer valid after the election ends. This diminishes the risk that decisions will censor lawful speech outside of election periods and make it harder for Brazil to impose its standards globally.

Nevertheless, under the same criteria, the description of how the system worked on the ground showcases the many problems and limitations of the Brazilian model. The distinction between facts and opinion can be particularly blurred in a polarized election context.⁸¹ The ideological aspects of elections can also thrust courts into the middle of heated election debates, impacting their ability to act neutrally. This politicization of courts can be exacerbated by open-ended standards like the ones existing in Brazil (i.e., empowering judges to remove content with “*facts known to be untrue*” or prevent the spread of “*gravely decontextualized*” information), as they provide different judges with ample discretion in an area where reasonable people may disagree on what constitutes adequate outcomes.

This concern is compounded by the expedited time frame required by Brazilian electoral law, which regularly leads to decisions being taken without affording the parties a proper right of defense and issued with little to no reasoning. The numerous courts involved in the dispute, along with the limited appeals and reversals, resulted in the coexistence of many inconsistent decisions and prevented the establishment of clear judicial standards. Most rulings do not cite a single case law or any other data source other than the initial complaint. This suggests that decisions are often based on initial impressions that are rarely reevaluated. This also provides empirical support for the idea that legal standards appear overly broad, depriving judges of proper guidance and allowing them to rely primarily on their instincts or prior knowledge when deciding cases.

⁸¹ According to Robert Post, factual statements claim general validity regardless of community standards, assuming it is possible to achieve convergence on the evaluation of truth or falsity, while statements of opinion claim validity grounded on certain community standards. Robert C Post, *The Constitutional Concept of Public Discourse: Outrageous Opinion, Democratic Deliberation, and Hustler Magazine v. Falwell*, HARV. LAW REV. 601, 660 (1990).

Another major shortcoming of the system is that it is excessively based on the notice-and-takedown regime created for copyright, which has already been shown to be defective at best.⁸² Altogether, Brazilian courts adjudicated 1,500 disputes and removed approximately 1,600 URLs from the internet during the 2018 elections—in a highly connected nation of 210 million inhabitants, that is nothing. While we do not have comprehensive information for the 2022 election period, it is unlikely that courts made a significant difference in the overall amount of disinformation circulating during that election, either. The system is also costly to maintain and is not designed to be scaled up to handle an increased number of cases. Placing the burden almost solely on political parties, politicians, and public prosecutors to monitor the avalanche of online speech is costly and ineffective in a world where information is publicly shared (e.g., Facebook, Instagram, YouTube, TikTok)⁸³. It will become increasingly untenable as interactions shift toward encrypted, peer-to-peer communication systems, such as WhatsApp, where there is no clear shared public sphere. Brazilian politicians attempted to include WhatsApp in the scope of the law by utilizing WhatsApp Web and other links; however, the system is not designed for this type of content.

c. Five more general takeaways for other jurisdictions

These, however, are more targeted observations that focus on improving the Brazilian system. Above all, we believe the Brazilian experience teaches five more general lessons to the broader world, which still struggles with problems surrounding content adjudication and widespread disinformation.

First, content moderation systems must clearly articulate their end goals, which the Brazilian system does not. A system focused on URLs and case-by-case adjudication might make sense if the ultimate goal is to shape the public discourse by having an official Court ruling that targets solely the most egregious types of disinformation—the ones that theoretically can discredit the whole system or shape the outcome of an election (e.g., that the Brazilian voting machines are easily hackable or that the one candidate is running a pedophile ring out of a pizza parlor).⁸⁴

However, suppose the goal is indeed to use the judiciary to discredit the most egregious falsehoods. In that case, judicial decisions should be carefully

⁸² See Sharon Bar-Ziv & Niva Elkin-Koren, *Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown*, 50 CONN REV 339 (2018).

⁸³ On the limitations of traditional speech-regulation rules to the internet, see Evelyn Douek, *Governing Online Speech: From "Posts-as-Trumps" to Proportionality and Probability*, 121 COLUM REV 759 (2021).

⁸⁴ For an analysis of the viral spread of the Pizzagate conspiracy, see Paul Bleakley, *Panic, Pizza and Mainstreaming the Alt-Right: A Social Media Analysis of Pizzagate and the Rise of the QAnon Conspiracy*, 71 CURR. SOCIOL. 509 (2023).

crafted and well-reasoned to help convince potential voters that the underlying claim is indeed false. Perhaps all adjudications should be conducted by judicial panels to enhance deliberation and credibility, and the Court should develop a strategy to disseminate its opinions to the broader public. The judicial system should also have discretion to rule on the cases that really matter, not spending precious energy and resources on minor cases of an obscure comment that will have little or no impact on the information environment. The Brazilian system lacks most of the characteristics that would enable authoritative court rulings to settle a given dispute, other than a general confidence in judges adjudicating the issue. That is not enough to establish credibility, especially in a world where parts of society increasingly see Courts themselves as political actors.

In contrast, if the goal is to diminish the overall amount of disinformation circulating in Brazil during an election period, then the system must be designed to operate at scale and in conjunction with the digital platforms' internal tools. In that case, a URL or profile-based notice and takedown system is playing a game of whack-a-mole that is impossible to win. For that system to work, Courts must refocus their attention on creating clear precedents and standards that platforms, candidates, and civil society can incorporate into their own content moderation efforts. Reforms in this area would better integrate civil society and digital platforms into the system, strengthening a coherent appeals process and allowing standards to emerge from caselaw in a more refined and objective manner. Without clearer standards, the impression of random decisions abridging the freedom of expression will undermine the legitimacy of court interventions on alleged disinformation. The Brazilian system sits between the two goals, and this limits its efficacy and amplifies criticism.

Second, experimentation in content moderation is possible and desirable—it will not break the internet. Platforms recurrently cry foul and argue that any weakening of speech-safe harbors will lead to the breakdown and demise of the open web.⁸⁵ The Brazilian example shows (yet again⁸⁶) that this is really not the case. Democratic societies will only be able to develop effective standards for online content moderation through experimentation, and it is

⁸⁵ See, for example, Adam Allington, *Would Reforming Section 230 Break the Internet?*, BLOOMBERG GOVERNMENT (June 21, 2021), <https://about.bgov.com/insights/news/would-reforming-section-230-break-the-internet-podcast/> as an example of the discussion on whether the reform of safe harbors for online speech would break the internet; see also Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans Sec. 230 Immunity*, 86 *FORDHAM REV* 401 (2017). (discussing these claims in more detail, and using theory to dismiss them).

⁸⁶ On the robust discussion and proposals to reform laws around online content moderation, see, for example, Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, *U CHI LEG. F* 45 (2020).

perfectly possible that different societies around the world will settle for different standards on what is accepted online and offline speech.⁸⁷

The decisions of Brazilian courts, with all their limitations, represent a significant experiment in attempting to create democratically acceptable limits for online speech based on real adjudication, thereby opening up space for a learning-by-doing exercise. For example, the new regulations issued between the first and second rounds of the 2022 elections were based on a general understanding that the system had gaps that enabled the systematic and repeated spread of fake news by organized groups. A solution was to make new remedies available for judges involved in these disputes. The same rules paid special attention to fake news aimed at undermining the integrity of the election process, as this type of disinformation may be particularly harmful to electoral legitimacy. These are novelties crafted through the learning process of the justice system, and they are necessary if societies are to implement effective remedies in complex and dynamic areas that are characterized by strong information asymmetries between parties.⁸⁸ Brazil's experience teaches that, in order to learn and develop new standards, it is necessary to experiment and adapt.

Third, a content moderation system must have strict protections to prevent a slippery slope from the outset, including a requirement for supermajorities in the development of such standards, as well as automatic sunset clauses and other similar provisions. The Brazilian system's creation was well-intended and relied on a set of tweaks to the case-by-case adjudication provisions initially established for television and radio advertisements. Over time, however, the system has been appropriated by an entrepreneurial Supreme Court Justice, with clear support of the majority of the Court, in a manner that challenges basic principles of due process and a more general commitment to freedom of expression. In particular, the concentration of authority in the Supreme Court (and, especially, in the Justice Rapporteur of the cases); the extensive investigations opened by the Supreme Court itself without a clear scope, timeline or end goal; the lack of procedural safeguards that protect defendants from abuse; and the lack of transparency in inquiries conducted by the Supreme Court; gradually eroded internal checks and balances in the system and opened room for arbitrariness. It is quickly pushing Brazil into a politically charged debate about the very boundaries of freedom of speech that extends well beyond the initial election period, which the Court was originally empowered to address. In practice, the Brazilian Supreme Court is increasingly becoming the ultimate content-moderation institution for the entire country, with the power to

⁸⁷ On the different societies developing different standards in terms of digital regulation, see Filippo Lancieri, *Digital Protectionism? Antitrust, Data Protection, and the EU/US Transatlantic Rift*, 7 J. ANTITRUST ENFORC. 27 (2019).

⁸⁸ Lancieri and Pereira Neto, *supra* note 23.

determine the exclusion of content, profiles, and even the suspension of access to digital platforms altogether.

This is an example of the slippery-slope outcome of content moderation that is so feared by constitutional scholars, one that justifies stricter speech protections that almost prevent any governmental or judicial intervention.⁸⁹ However, this outcome is not unavoidable. The Brazilian case builds on the numerous previous weaknesses of the Brazilian Supreme Court adjudication system, which is overly concentrated on Justice Rapporteurs, who have broad powers to issue individual decisions, in a manner that risks being arbitrary and challenges basic due process notions from the outset⁹⁰. It can be adjusted through specific reforms that reinstate the boundaries between prosecution and adjudication, require decisions in panels—potentially requiring super-majorities for key cases—and a more general decentralization of power within the system, among other ideas that are beyond the scope of this article.

Fourth, any content-based moderation system becomes part of the information environment and needs to consider the reactions it generates and the impacts on its own legitimacy. In the Brazilian case, the courts' action has generated censorship concerns. This reaction has been particularly intense among right-wing parties and their supporters,⁹¹ who have criticized the courts for restricting freedom of expression and showing an unfair political bias when taking down content.⁹²

This has created a negative feedback loop: The very act of taking down content brings more attention to that information in a variation of the so-called Streisand Effect.⁹³ The visibility of the take-down order sparks an anti-censorship movement, attacking the Courts and the moderation system.⁹⁴ This

⁸⁹ Kosseff, *supra* note 44.

⁹⁰ See the many works of Diego Werneck Arguelhes outlining the many shortcomings of the Brazilian Supreme Court model. For example, Diego Werneck Arguelhes & Leandro Molhano Ribeiro, *Ministrocracia: O Supremo Tribunal Individual e o Processo Democrático Brasileiro*, 37 NOVOS ESTUD. CEBRAP 13 (2018).

⁹¹ See, for example, FGV's study identifying that "right wing leverages Elon Musk situation on social media to drive narratives of 'censorship' and 'dictatorship' in Brazil". Available at FGV Cemi, *Direita Mobiliza Caso Elon Musk Nas Redes Sociais Para Legitimar Denúncias de "Censura" e "Ditadura" No Brasil | Mídia e Democracia*, (Nov. 4, 2024), <https://midiademocracia.fgv.br/estudos/direita-mobiliza-caso-elon-musk-nas-redes-sociais-para-legitimar-denuncias-de-censura-e-0>.

⁹² To our knowledge, there are no systematic studies evaluating whether court decisions are being applied disproportionately to right-wing or left-wing posts (clearly a point that deserves further research). However, public reactions against the removal of content have been clearly more intense and frequent from right-wing groups.

⁹³ See <https://www.britannica.com/topic/Streisand-effect>

⁹⁴ As pointed out by Tim Wu, the flooding the public sphere with disinformation is usually protected by an anti-censorship principle, under the right of free speech. Because this

contentious environment increases the incentives of those who feel censored to flood the public sphere with additional and more extreme disinformation in favor of their agenda: on one hand, if the courts strike down pieces of the additional content, it will draw more attention to it and fuel the anti-censorship movement; on the other hand, to the extent that courts do not take down the content (as we saw, only a tiny fraction of the content is taken down at any given time), it favors the group flooding the public sphere. The net effect may increase the amount of disinformation in the public sphere, requiring courts to respond, starting the loop again, escalating the arms race.⁹⁵

This apparent paradox illustrates the complexity of the task. Because the courts are themselves part of the information environment in which they intervene, their decisions to take down content generate immediate reactions, becoming part of the public debate in a democratic environment, fueling information strategies of different groups. Thus, no institution responsible for a content-based moderation system can act as if they are simply a neutral umpire without consideration for its own sources of democratic legitimacy.

Two types of strategies may help mitigate this problem. First, one must acknowledge that pure content-based moderation systems have limitations and courts (or any other independent institution implementing them) must develop a focused strategy, using their energy and political capital wisely, to target the organized spread of disinformation, coordinated attacks to institutions and integrity of elections and the funding of disinformation campaigns—this means abandoning the “retail” effort to evaluate content one-by-one. Second, these authorities must also focus on developing content-neutral strategies to deal with disinformation, such as: targeting robot-aided flooding, increasing transparency regarding paid social media advertisement, user control over curation/recommendation algorithms, limitations on group sizes, and mass forwarding of content in messaging apps, among many others. These can be considered architectural strategies aimed at reducing the impact and

guarantee protects speech broadly, it may be used to shield speech generated to suffocate other speech from court intervention and regulations. Thus, illiberal actors can use freedom of speech and anti-censorship principles as part of their broader strategy to control the public sphere. See Tim Wu, *Is the First Amendment Obsolete?*, KN. FIRST AMEND. INSITUTE COLUMBIA (2017),

<https://knightcolumbia.org/sites/default/files/content/Emerging%20Threats%20Tim%20Wu%20Is%20the%20First%20Amendment%20Obsolete.pdf>.

⁹⁵ For a more detailed discussion of this feedback loop, see Caio Mario Pereira Neto, *Brazil's Efforts To Address Election Disinformation Illustrate the Difficulties of Protecting the Marketplace of Ideas*, PROMARKET (June 27, 2025),

<https://www.promarket.org/2025/06/27/brazils-efforts-to-address-election-disinformation-illustrate-the-difficulties-of-protecting-the-marketplace-of-ideas/>.

dissemination of disinformation without incurring allegations of viewpoint discrimination.⁹⁶

Fifth, we need to develop a new system of international rules and limits for extraterritorial online decisions. The US Government openly accused Brazilian courts of violating US laws regarding free speech by requiring the takedown of accounts based in the US. In response, it imposed 50% tariffs on international trade as a retaliation measure. As briefly discussed above, taken to the letter, their underlying accusation is correct: Brazilian court orders effectively ordered the takedown of accounts belonging to Brazilian citizens residing in the US, so technically under the purview of US laws. However, these accounts also directly targeted Brazilian audiences and attacked Brazilian courts and democratic institutions. Consequently, their content had evident repercussions in Brazil, impacting Brazilian sovereignty. Refusing to address problems because of the location of the content creator would create massive enforcement gaps and open room for regulatory arbitrage. A content provider might simply move to another jurisdiction to evade enforcement from national regulators without punishment.

The Brazilian experience gives new urgency to the quest to reassess the limits of international law and governance in the context of the internet⁹⁷. In particular, the current geopolitical scenario of increasing regulatory fragmentation highlights the need to develop at least some guidelines on the acceptable limits of extraterritorial decisions and what types of practices cross reasonable boundaries. It is not normal, but it may be increasingly inevitable, that large jurisdictions such as the US, the EU, and Brazil will accuse each other of meddling in their internal affairs because of the way they regulate digital platforms. Societies must create mechanisms to mitigate these conflicts, lest they quickly escalate out of control⁹⁸.

The location of the primary audience of the content is clearly one element to be considered in this regard and the decision to locate the content in another jurisdiction to evade local regulations may be another.

Finally, in designing any system of content moderation, and using new regulatory tools, it is important to keep in mind that large digital platforms are reaching a level of economic and political power that enables them to challenge (or to serve as a vehicle for others to challenge) the democratic institutions of

⁹⁶ *Id.*

⁹⁷ On broader questions around sovereignty and internet regulation, see Anupam Chander & Haochen Sun, *Sovereignty 2.0*, 55 VAND J TRANSNATL L 283 (2022); Anupam Chander, *When the Digital Services Act Goes Global*, 38 BERKELEY TECH LJ 1067 (2023). On the limits of international law, see Evelyn Douek, *The Limits of International Law in Content Moderation*, 6 UC IRVINE J INTL TRANSNATL COMP L 37 (2021).

⁹⁸ On the risks of arbitrage and extraterritoriality of legal rules online, see Lancieri, Edelson, and Bechtold, *supra* note 9.

almost any jurisdiction. Scholars have long been warning societies about the growing economic and political power of digital platforms.⁹⁹ In this context, democratic societies must experiment with new institutional models that enable them to regain the upper hand and rebalance this increasingly tilted playing field. It will take time and a steep learning curve to reach a new desired equilibrium. Yet, the hands-off approach to digital regulation of content moderation, adopted by many countries so far, has not shown positive results and will only allow these giant digital platforms to continue growing unabated, eroding fundamental notions of democratic self-determination.

V. Conclusion

It has been six years since we started working on this article. In the meantime, the discussion around fake news and disinformation has certainly evolved, but it has never abated. Our hyperconnected, digital platform-mediated public sphere has reconfigured the boundaries of freedom of expression, public participation, and even the difference between facts and opinions.

As societies develop much-needed governance models and infrastructure for this new world,¹⁰⁰ they must learn from international experimentation. We hope that, in doing so, they also learn the good and bad lessons from Brazil's continued use of Courts as the key agents in the fight against disinformation, both within and increasingly outside election periods.

⁹⁹ JULIE E. COHEN, BETWEEN TRUTH AND POWER (2019); BRADFORD, *supra* note 23; Zephyr Teachout & Lina M. Khan, *Market Structure and Political Law: A Taxonomy of Power*, 9 DUKE J CONST PUB POL 37 (2014); Julie E Cohen, *Oligarchy, State, and Cryptopia*, Forthcoming FORDHAM REV (2026), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5171050.

¹⁰⁰ On the need to reimagine public spheres and infrastructures, see the foundational work of COHEN, *supra* note 98; Julie E Cohen, *Infrastructuring the Digital Public Sphere*, 25 YALE JL TECH 1 (2023).

Appendix I: Measuring associations in the takedown orders for the 2018 election decisions

Table 1 provides summary statistics. Data available for each decision tracks citations of nine relevant articles of law, citations of court cases, how judges utilized evidence for standards of truth, characteristics of individual judges, and characteristics of each decision. We have:

- i. Grouped up mentions to legal articles of the Brazilian Constitution and related regulations that promote freedom of expression¹⁰¹ in a binomial that is assigned to one whenever any of these articles are cited;
- ii. Created a binomial variable for procedural articles¹⁰² that generally empower the courts to act and order the removal of online content;
- iii. Grouped articles that stipulate prohibited electoral advertising, facts known to be untrue, and the electoral crime of libel, defamation, and sharing of false information;¹⁰³ and finally
- iv. The “Standards of Truth” category refers to how judges decide whether the underlying content involved a “fact known to be untrue,” which is the official legal standard. We divided cases under the categories “not considered” (when the judge does cite the legal standard “facts known to be untrue”), no additional sources (meaning that the judge does not cite any additional source when discussing whether the alleged content promotes a “fact known to be untrue”), the use of non-official sources such as press reports or other material in discussing whether there is a “fact known be untrue” in the online content, and the use of official government sources such as police report or others.

¹⁰¹ Federal Constitution Articles 5 and 2020, Resolution #23551 Article 33.

¹⁰² Resolution #23551 Article 6, Law #9504 Articles 57 and 58.

¹⁰³ Resolution #23551 Articles 17, 22, and 85, and Article 325 of the Electoral Code.

Table 1: Descriptive Statistics of Decisions

	Avg.	Std. Dev.	Variance	N
Content Removal	0.51	0.50	0.25	2186
Freedom of Expression	0.54	0.50	0.25	2186
Procedural Matters	0.57	0.50	0.25	2186
Electoral Advertisement	0.38	0.49	0.24	2186
Court Case Citation	0.47	0.50	0.25	2186
Standards of Truth: Not Considered	0.42	0.49	0.24	2186
Standards of Truth: No Additional Sources	0.40	0.49	0.24	2186
Standards of Truth: Non-official Sources	0.09	0.29	0.08	2186
Standards of Truth: Official Sources	0.09	0.28	0.08	2186
Federal Judge	0.17	0.37	0.14	2186
State Judge	0.63	0.48	0.23	2186
Appointed Lawyer	0.20	0.40	0.16	2186
Female Judge	0.21	0.41	0.17	2186
Interim Ruling	0.53	0.50	0.25	2186
Appeal	0.09	0.28	0.08	2186
District Court Ruling	0.38	0.49	0.24	2186
State GDP per Capita	32686.31	14967.98	224040420.92	2186
State Population	26241.71	50597.29	2560086154.88	2186

Results are presented in Models 1 through 3, which estimate the probability of content removal using a logit regression. Coefficients are exponentiated to facilitate interpretation in odds ratios. Model 1 shows the variables grouping articles under three different binomials, while Model 2 emphasizes categories of standards of truth compared to not utilizing the term “facts known to be untrue.” Model 3 offers a full specification. All models present state judges and appointed lawyers compared to the baseline of federal judges, women judges compared to the baseline of male judges, appeals and district court rulings compared to the baseline of interim rulings, and the GDP per capita and population estimates for the state in which a given court is located. For the Supreme Electoral Court, estimates for the country were used instead.

Table 2: Model Results

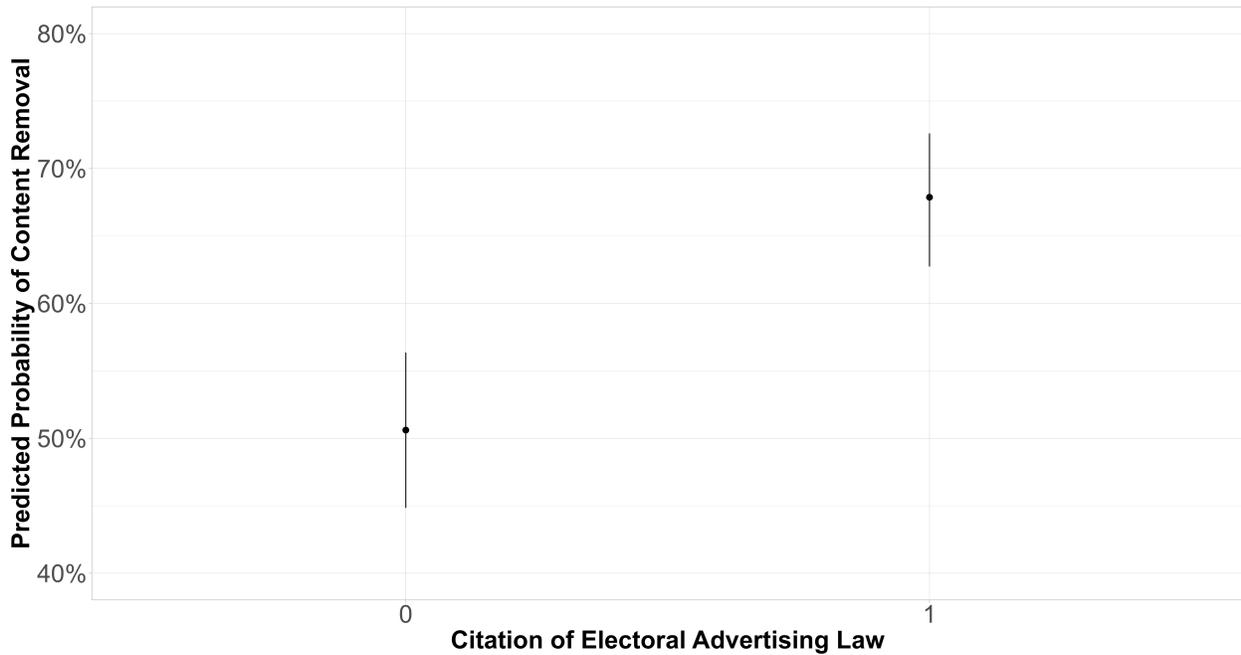
	Model 1	Model 2	Model 3
Constant	4.702 (5.061)	2.938 (3.064)	5.203 (5.625)
Freedom of Expression	0.848+ (0.081)		0.858 (0.082)
Procedural Matters	1.107 (0.103)		1.106 (0.104)
Electoral Advertising	2.061*** (0.198)		2.055*** (0.199)
Court Case Citation	0.447*** (0.042)		0.446*** (0.043)
Standards of Truth: No Additional Sources		1.003 (0.096)	0.991 (0.099)
Standards of Truth: Non-official Sources		0.860 (0.138)	0.846 (0.142)
Standards of Truth: Official Sources		1.570** (0.258)	1.542* (0.263)
State Judge	1.068 (0.137)	1.259+ (0.155)	1.058 (0.136)
Appointed Lawyer	1.725*** (0.262)	1.693*** (0.248)	1.716*** (0.262)
Female Judge	0.926 (0.103)	0.954 (0.104)	0.929 (0.104)
Appeal	0.775 (0.128)	0.638** (0.103)	0.761+ (0.126)
District Court Ruling	1.077 (0.105)	0.926 (0.085)	1.071 (0.104)
State GDP per Capita (logged)	1.097 (0.124)	1.166 (0.129)	1.087 (0.124)
State Population (logged)	0.856*** (0.030)	0.838*** (0.028)	0.856*** (0.031)
Num. Obs.	2186	2186	2186
AIC	2876.2	2996.7	2873.2
BIC	2944.5	3059.3	2958.6
Log. Lik.	-1426.117	-1487.342	-1421.607
F	14.510	5.314	11.910
RMSE	0.48	0.49	0.48

Model 1 Model 2 Model 3

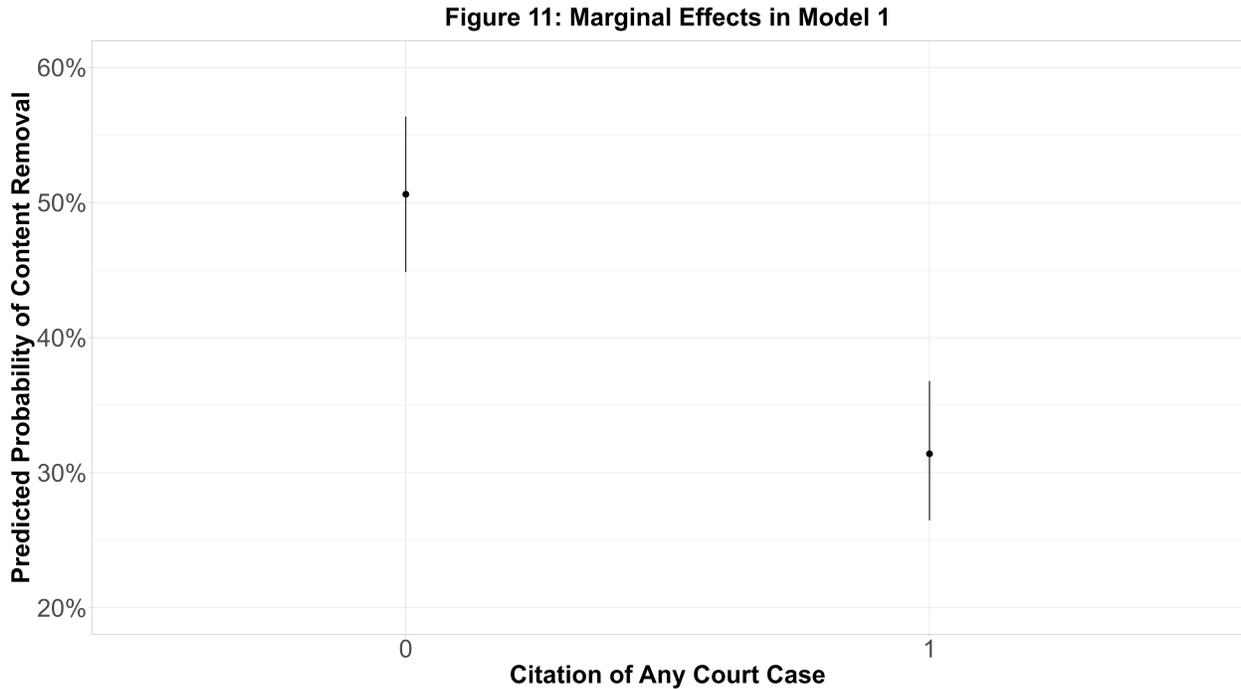
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Among the groups of articles cited, only articles on electoral advertising are consistently significant. In Model 1, citing any relevant article on electoral advertising increases the odds of content removal by 2.061 times. This effect is significant at the .001 level. With the specification in Model 3, its effect marginally shifts to an increase in odds by 2.005 times, maintaining the significance level. The estimated marginal effects for citing electoral advertising law in Model 1 can be visualized in Figure 10. The estimated probability of removal moves from the baseline of 50.6% to 67.8% when decisions cite electoral advertising law.

Figure 10: Marginal Effects in Model 1



While the variable for freedom of expression articles decreases the odds in the expected direction, its coefficient is not significant in any model. Most interestingly, citing any court case substantially reduces the odds of removal by 0.447 times in Model 1, which is significant at the .001 level, with magnitude and significance practically unchanged in Model 3. The estimated marginal effects for citing any court case in Model 1 are shown in Figure 11. The estimated probability of removal decreases from the baseline of 50.6% to 31.3%.



Compared to cases that do not mention “facts known to be untrue,” using official government sources as standards of truth increases the odds of removal by 1.570 times in Model 2 and 1.542 times in Model 3, which are significant at the .01 and .05 levels, respectively. Interestingly, compared to Federal Judges, appointed lawyers present an increase in odds of content removal by 1.725 times in Model 1, which is significant at the .001 level. Other models present coefficients with similar magnitude and significance.

Overall, significant indicators of increased odds of content removal include citations of electoral advertisement law, the use of official government sources as standards of truth, and the appointment of lawyers compared to Federal Judges. Surprisingly, citing the law on freedom of expression does not significantly decrease the odds of content removal. Instead, citing any court case throughout the ruling seems key to lowering the odds of content removal.

While the baseline takedown rate is 50.6%, the odds can be even higher given other relatively common circumstances. For example, electoral advertising law is cited in 38% of decisions, and appointed lawyers are involved in 20% of decisions. Meanwhile, a primary mitigator for content removal, citing any court case, is seen in 47% of the decisions.