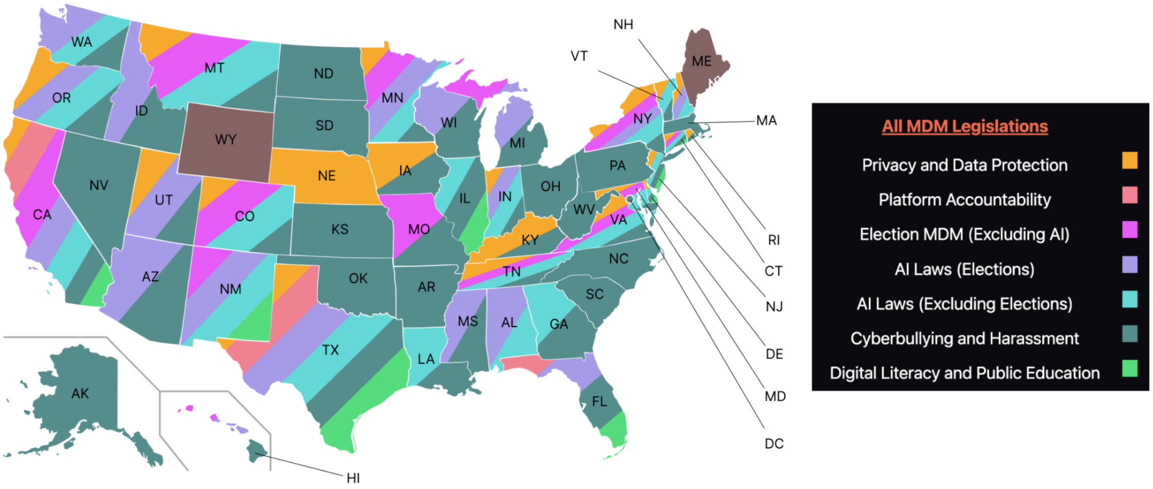
NYM

Protecting Free Speech While Combating Digital Disinformation Is Harder Than Ever

Research and Writing by Resident Researcher, Dorna



Key Takeaways

1. Section 230 in the US: Is it Protecting Free Speech or the Spread of MDM?

Section 230 of the Communications Decency Act (1996) protects platforms from liability for user content, but many critics point to its current limitations. States like California push for transparency, while Texas and Florida are focused on preventing platform censorship. The divide in focus mirrors broader binaries: public health vs. election integrity and privacy vs. censorship. Marginalized communities, like Black and Latino voters, are disproportionately targeted by MDM, complicating the already controversial landscape of state vs. federal control.

2. State-Level MDM Regulation Intensifies

California leads other states in regulating MDM with laws spanning all seven identified categories. Texas follows closely with six categories, while New York and Connecticut round out the top with regulations covering five categories. This reveals a growing trend of intense regulation at the state level. However, our data does not provide strong enough evidence to claim a positive correlation between Democratic-leaning states and the number of MDM law categories. The regression analysis shows a coefficient of 1.48, suggesting that Democratic states may tend to have more MDM regulations. However, with a p-value of 0.138, the result is not statistically significant.

3. The Absence of MDM Regulation in Wyoming and Maine

Wyoming and Maine are the only US states without significant state-level laws addressing MDM in public discourse or on digital platforms. This regulatory gap leaves both states particularly vulnerable to the unchecked spread of harmful content, especially during critical moments like elections or public health crises.

4. MDM Worldwide: Big Tech vs. Governments Regulation Showdown

The escalating tension between Big Tech and governments over content regulations is reshaping the internet. Brazil's nationwide ban on X, the arrest of Telegram's founder in France, and Russia and Turkey banning Discord are all signs of growing clashes. These incidents highlight the need for a new debate on the merits of net neutrality and whether the free flow of data is still viable in today's heavily monitored digital world.

5. The Four Global Models of MDM Regulation: US vs. China at Opposite Poles

MDM regulation spans four models worldwide, with the US and China as extreme opposites. The US adopts a laissez-faire model, relying on free speech protections and Section 230 for platform immunity, while China enforces a strict liability model that demands proactive censorship. Hybrid models like the conditional immunity approach of the EU, UK, and India, as well as the repressive criminalization model seen in Russia, Belarus, and Egypt, are also included.

6. The False Binary: “Good” vs. “Bad” Governments in MDM Regulation

It's not just authoritarian regimes that are tightening control. Even democratic countries like Germany have adopted harsh regulations like those of NetzDG, putting them in the same category as China for proactive content removal. The critical difference? While China focuses on broad state censorship, Germany's approach targets illegal content but with steep fines, often incentivizing over-censorship. Authoritarian states like Russia and Venezuela are now referencing Germany's model as a blueprint to justify crackdowns on free speech.

7. Russia vs. China: MDM Crackdowns with Different Flavors

Though Russia and China are often lumped together in discussions about censorship, their tactics are distinct. China focuses on state-controlled, real-time filtering through its infamous “Great Firewall,” effectively censoring at the source. Russia, on the other hand, leans on punitive laws – criminalizing dissent and targeting journalists – using misinformation laws as a tool for strategic repression, particularly around national security.

Full research database here

Net Neutrality Revisited

The debate around freedom of information and misinformation is evolving rapidly, but not necessarily in the ways we expect. Governments, even in so-called democracies, are increasingly caught between allowing free speech and protecting public order from harmful misinformation, disinformation, and malinformation (MDM). The consequences of failing to find a balance can be dramatic, as we've seen with the spread of misinformation during the COVID-19 pandemic and the influence of disinformation and false narratives in US elections.

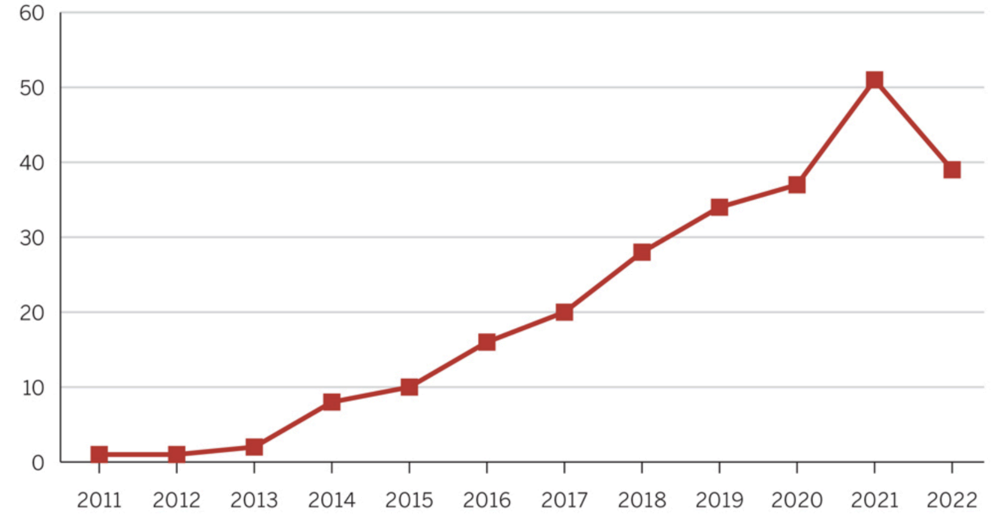
A recent example that caught the world's attention was Brazil's bold move to block X nationwide, targeting its role in spreading political misinformation. Under the leadership of Supreme Court Justice Alexandre de Moraes, this decision followed an extended legal conflict with Elon Musk, who refused to comply with local demands to moderate content deemed harmful to Brazil's democratic integrity. The move raised alarms when Brazil temporarily banned VPNs, a privacy tool that roughly 37% of Brazilians used to bypass the block. Though some argue that the overreach compromised civil liberties, it also reflects a rising global concern over the unchecked power of foreign big tech companies to influence local public debate.

This is not an isolated case. Another tech giant, Telegram, has faced controversy for its role in enabling criminal activity despite being framed as a beacon of free speech. Following his recent and highly publicized arrest in France, Telegram founder Pavel Durov revealed that the platform has been cooperating with authorities in countries like Brazil and India. He disclosed that Telegram had provided user data in compliance with legal requests tied to severe crimes such as terrorism and organized crime. However, knowing that such legal cases rarely reveal full details to the public, it's naive to assume that criminal activity is the sole reason behind Telegram's ongoing legal challenges. This raises more questions about the real motivations at play, especially considering that the president of Signal, a private messaging protocol and application, also resides in France but has faced no similar legal persecution. The contrast between the two cases suggests deeper issues may be behind government scrutiny of platforms like Telegram. It also raises questions about whether Telegram, compared to applications like Signal, truly offers the level of privacy its recent marketing campaigns have claimed.

This growing pattern of government intervention in online platforms illustrates a broader trend. More recently, Turkey and Russia joined the fray by banning Discord after it was linked to severe allegations of child abuse and cyberbullying. Discord has faced scrutiny for hosting illegal content, from child exploitation to cybercrime, prompting sudden government intervention. This marks the global nature of the problem: as more platforms are targeted, the line between legitimate governance and excessive control grows increasingly thin. **The concepts of misinformation and censorship have become weapons in the current battles between governments, big tech, and their geopolitical alliances, with each side wielding them to protect their interests and influence public opinion.**

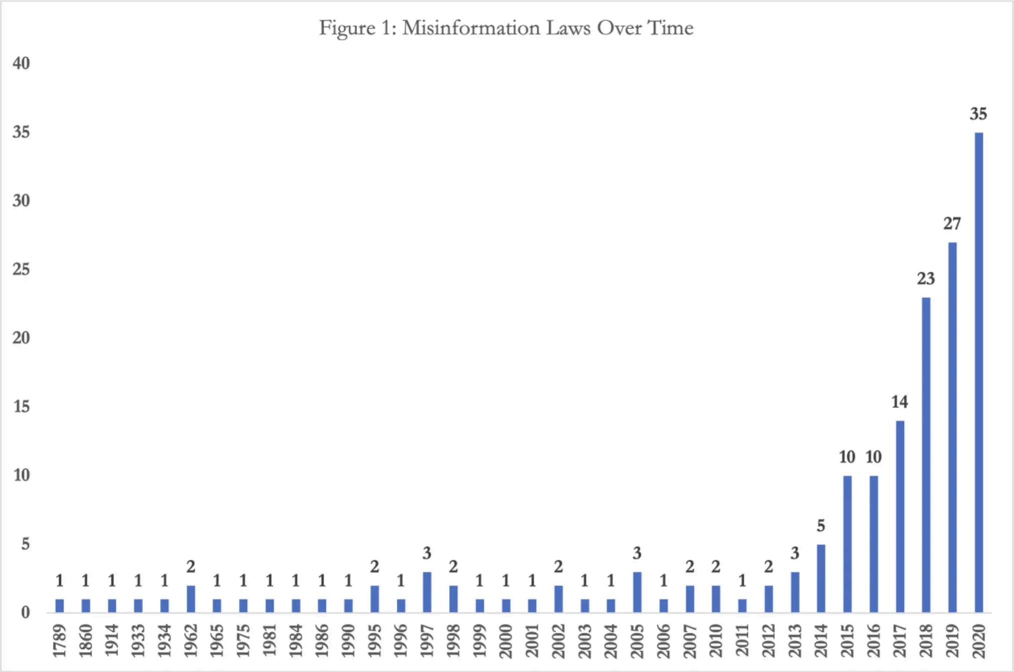
In surveillance-dense regions, where governments tightly monitor public discourse, the debate is more nuanced and often dangerous. Governments in these areas justify content regulation by framing it as protecting public order, but the risks of such policies are clear. When authorities are empowered to decide what constitutes "illegal" content, the line between legitimate regulation and oppressive censorship quickly becomes blurred. In particular, ordinary citizens, journalists, and dissidents are the most vulnerable. They risk persecution for simply voicing critiques of the government or challenging official narratives. In such contexts, "illegal content," including MDM, becomes a catch-all term that allows authoritarian governments to quash dissent and control the political narrative.

Number of Journalists Imprisoned on MDM Charges



SOURCE: Committee to Protect Journalism, Imprisoned Journalists Annual Data

In contrast, in more democratic countries, the issue redirects to how MDM manipulates public interest, threatening public health, political agency, and election integrity. Many countries are grappling with how to regulate harmful content – without inadvertently slipping into censorship. Even in democratic countries, vague wording in MDM regulations can lead to unintended censorship. Take Germany's 2017 Network Enforcement Act (NetzDG) as an example. Initially designed to combat hate speech online, critics have argued that its broad and unclear definitions risk overreach. Heavy fines imposed on social media platforms could incentivize companies to remove more content than necessary to avoid penalties, potentially stifling political speech. The law lacks sufficient judicial oversight or an appeals process, making it harder for users to contest blocked content. Moreover, this law has raised concerns internationally, as authoritarian regimes could use it as a model to justify suppressing online dissent under the guise of regulating harmful content. Countries like Russia, the Philippines, and Singapore have referenced NetzDG while drafting or enforcing laws that similarly target online speech, but often in ways that can crack down on political opposition or free speech more broadly.

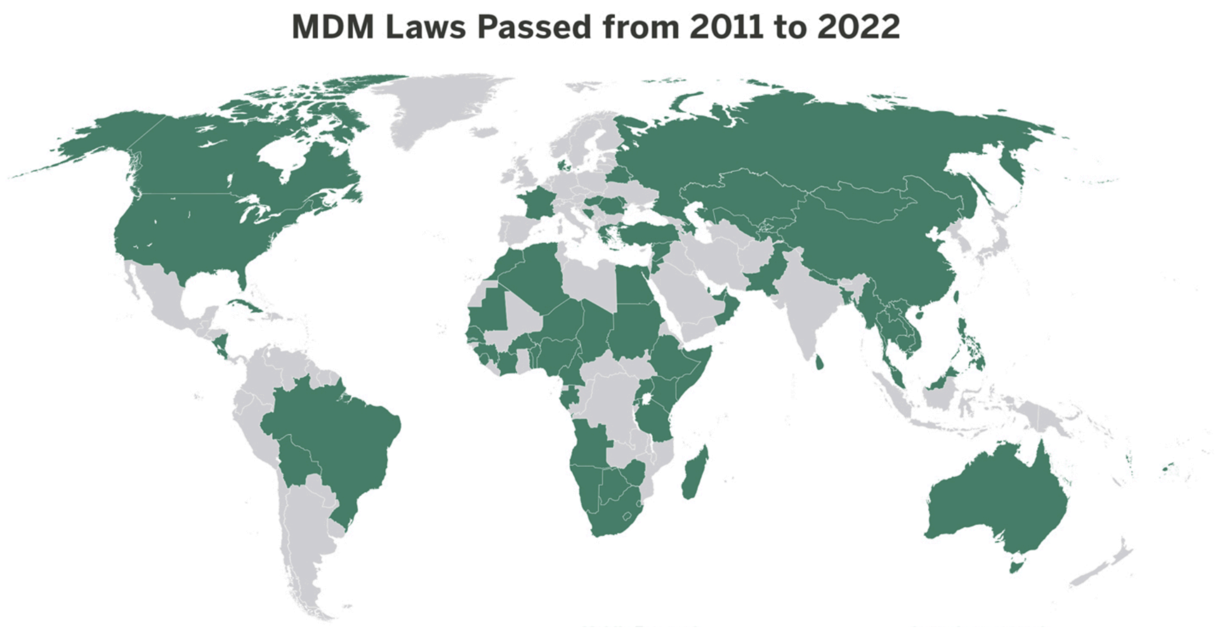


Since 2019, governments have implemented, amended, or proposed at least 62 laws to address misinformation.

Source: Bulletin

Therefore, it's not simply a battle between "good" and "bad" governments; every state – regardless of its democratic status – faces the complex challenge of balancing free expression with the need to protect its citizens from disinformation's corrosive effects. In 2016, the US saw widespread disinformation campaigns aimed at undermining electoral trust, and in Brazil's 2022 election, online falsehoods contributed to violent unrest. Or consider public health during COVID-19: misinformation about vaccines, masking, and treatments caused confusion and slowed down efforts to control the virus. Conversely, the subsequent clampdown on misinformation also shut down any meaningful discussion about vaccine side effects and effectiveness, limiting public access to important, balanced health information.

This leads us to the issue of “net neutrality,” the principle that internet service providers should treat all data equally without discrimination or charging differently per user, content, or platform. Once hailed as a manifesto for free and open access to information, it now seems out of step with the realities of modern digital society. Initially celebrated for preventing the throttling of internet services, net neutrality as a principle hasn’t been adequately updated to deal with the mass manipulation of public discourse. The debate has suffered from a lack of profound intellectual and scholarly engagement in recent years. With increased digital surveillance, the spread of misinformation, and the impact on public opinion and political agency, it’s clear that old debates on free information flow need serious updating.



Source: CIMA

The overwhelming amount of data people consume today, whether from social media, news websites, or even apps, directly influences social behavior, political decisions, and the very fabric of democracy. As more data becomes available, often unchecked and easily manipulated, governments are increasingly pushed to take a role in moderating or even controlling what is shared: Where should governments draw the line? Should regulation be used to protect the public from dangerous misinformation, or could it quickly become a tool for excessive state control over free speech? At what point does content “moderation” slip into censorship?

Complicating this issue further, private companies are often responsible for managing content. The current approach of many Western governments to safeguarding freedom of expression online mandates that internet platforms determine what content is acceptable, display these policies in their terms of service and apply them consistently. However, this model has a critical flaw: there is little regulatory oversight governing what companies include in their terms of service and how they implement these rules. This lack of transparency and accountability means that platforms can moderate content based on internal interests, which fails to stop the spread of misinformation and raises concerns about selective enforcement and potential censorship.

Models of Online Content Regulation: A Global Overview

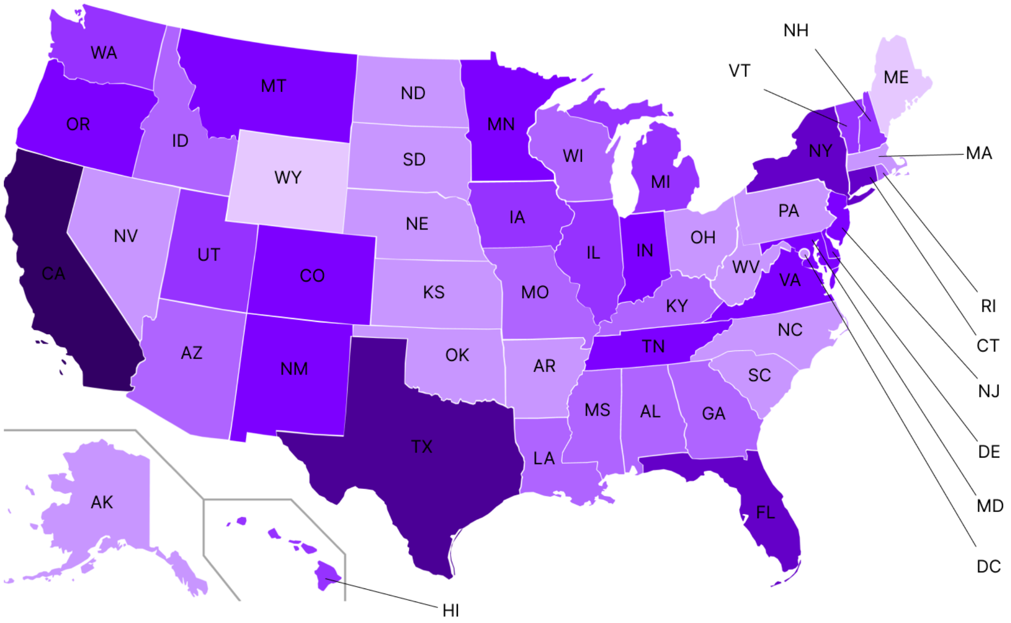
Online content regulation worldwide follows a spectrum of approaches, with the United States and China representing two starkly opposite poles. The United States leans toward a laissez-faire model, mainly avoiding direct government intervention in online speech. At the same time, China maintains one of the strictest regimes, enforcing strict liability where platforms are required to actively monitor and censor content.

Between these two extremes lie countries like India, the UK, and many European countries that follow hybrid models of conditional immunity or repressive criminalization. It is important to note that there is no universally agreed-upon solution to managing MDM. The US model will be discussed in depth in the subsequent section, particularly focusing on Section 230.

1. The Laissez-faire (Broad Immunity) Model: United States

The US model is grounded in the First Amendment, which offers robust protections for free speech, resulting in one of the most permissive regulatory environments for online content globally. Under Section 230 of the *Communications Decency Act (1996)*, platforms are granted broad immunity from liability for user-generated content, except in limited cases such as child pornography and intellectual property violations. This hands-off approach has led to platforms like Facebook and Twitter enforcing their own community guidelines, which are sometimes more restrictive than US law, particularly around hate speech.

This model is at the center of heated domestic debates, with increasing calls for reform. Bipartisan discussions in Congress revolve around reducing the scope of Section 230, with Republicans focusing on alleged political censorship and Democrats pushing for stricter moderation of harmful content. These debates stress a fundamental tension in US regulation: balancing preserving free expression while curbing online harms such as disinformation, especially in critical contexts like public health and elections. Meanwhile, states have intervened and passed laws to control the spread of MDM, creating a patchwork of legislative coverage across the country. These state-level efforts vary widely, from targeting election misinformation to addressing AI-generated content, reflecting each state's unique priorities and concerns.



Legislative Coverage

2. Conditional Immunity Model: European Union, United Kingdom, India

Many countries, particularly in Europe and Asia, have adopted conditional immunity models, which require platforms to remove unlawful content upon receiving notice. This middle-ground approach balances free expression with regulatory intervention, often focusing on specific harmful content like terrorism, hate speech, or misinformation.

- **European Union**: The *Digital Services Act* (DSA), which came into effect in 2023, sets a new standard in online content regulation in Europe. The DSA mandates platforms to remove illegal content, such as terrorism-related posts but also introduces mechanisms to regulate hate speech and misinformation more effectively. Before this, platforms voluntarily adhered to codes of conduct, such as the *EU's Code of Conduct on countering illegal hate speech*. However, the DSA now makes this obligatory, placing new accountability measures on platforms.

- **United Kingdom**: The UK passed the *Online Safety Bill* in the autumn of 2023, introducing stringent measures for regulating harmful content. Ofcom, the UK's communication watchdog, has significant powers to fine companies that fail to remove illegal content. This proposed legislation marks a shift toward a more proactive stance, edging closer to strict liability models like that of Germany. While traditionally strong on free expression, the UK's approach is more favorable to defamation claims, making it easier for individuals to sue for libel than the US. Interestingly, according to Ofcom's 2023 report on news consumption in the UK, trustworthy information websites in the UK attract around two billion visits per month, while false information websites receive only about 14 million visits — a significant 140:1 ratio.

- **India**: India's online content regulation is based on the *Information Technology Act* and its *Intermediary Guidelines* 2021. Platforms must remove content that violates national security or disrupts public order, failing which they lose their immunity from liability. While influenced by US and British legal traditions, India's regulatory framework leans toward conditional immunity, giving the government significant power to guide platforms on permissible content.

These countries demonstrate how conditional immunity models offer flexibility. Platforms must remove harmful content upon notice but are not required to actively monitor it unless a complaint is received. This model balances safety and free expression, though critics warn that stringent deadlines can force over-censorship.

3. Strict Liability Model: China and Germany

Countries like China and Germany exemplify the strict liability model, though the degree of censorship and government oversight differs significantly.

- **China**: In China, platforms must monitor, filter, and remove content proactively. The state enforces draconian measures, including the famous *Great Firewall*, to control online information. Content deemed harmful to state security, insulting to the government, or disrespectful to national symbols is quickly removed. Examples of this are numerous, such as scrubbing all mentions of tennis player Peng Shuai's allegations against a top official. China's regime is all-encompassing and permits minor to no room for dissent or freedom of expression.

- **Germany**: Germany's *NetzDG* law, introduced in 2018, also requires platforms to quickly remove illegal content such as hate speech and terrorist propaganda. However, unlike China, Germany's approach focuses on unlawful content rather than broader state censorship. Platforms must remove content within 24 hours after receiving a notice, with heavy fines for non-compliance. NetzDG also covers defamation, along with hate speech and terrorist propaganda, imposing strict removal deadlines on platforms. The intensity of censorship is significantly lower than in China, but the strict timeframes and steep fines force platforms to err on the side of caution, sometimes removing content that is not actually illegal.

While both countries employ strict liability models, Germany's approach is tempered by a focus on transparency and adherence to legal norms, while China's is more aligned with authoritarian control.

4. Repressive Criminalization Model: Russia, Egypt, Belarus, and Beyond

A growing number of countries, especially in more authoritarian regimes, have adopted repressive criminalization models. These laws hold platforms accountable and criminalize individual users for creating or sharing false or harmful content.

- **Russia**: Russia's expansion of its *fake news* law in 2022, in the context of its invasion of Ukraine, provides an example of how misinformation laws can be used to suppress dissent. Under this law, criticism of government bodies, including the military, can result in criminal charges, with journalists being arrested for reporting on the war. This law has heightened fears of self-censorship among Russian journalists and undermined the role of independent media.

- **Belarus**: Belarus has also implemented aggressive measures, amending its media laws to target “fake news.” Media organizations must register with the government, and all comments or posts on public forums must identify the author. This ensures that dissenting voices can be easily tracked and punished, further chilling freedom of expression.

- **Egypt**: In Egypt, journalists can be imprisoned for spreading what the state deems false information. The government’s broad definitions of fake news allow for the arbitrary detention of reporters and dissenters, with at least 19 journalists jailed under MDM-related charges in 2018 alone.

Comparing Russia with China, while both countries enforce strict censorship, China emphasizes real-time, state-controlled filtering and platform accountability. In contrast, Russia focuses more on punitive laws for speech that challenge the state narrative, using laws strategically, especially in the context of national security or military operations. These nuanced differences suggest that China’s model leans more toward broad, proactive censorship, while Russia’s model incorporates more of a targeted criminalization of dissent. Some other countries that have implemented repressive criminalization laws include Iran, Turkey, Thailand, Cambodia, and Kazakhstan. These laws are often framed as national security measures but are primarily used to silence opposition voices and stifle independent media.

Globally, online content regulation is highly diverse, with different countries adopting various models based on political, cultural, and social contexts. A recent global study covering 188 countries identified that only 35 nations, primarily in Europe, have enacted specific laws to combat disinformation. Countries like Ethiopia and Mauritania in Africa, along with India, Singapore, and Pakistan in Asia, stand out as regions where significant strides have been made to address MDM through regulation. However, 38 countries, including many in South America, have no regulation regarding MDM, instead relying on outdated laws or penal codes to address the spread of false information.

The American Approach to MDM Regulation and Section 230

The US approach to regulating online speech and misinformation has long been shaped by its deep-rooted commitment to free expression. This has resulted in a broad immunity framework under Section 230 of the Communications Decency Act (CDA), passed in 1996. Commonly referred to as CDA 230, this law essentially shields platforms like Facebook, Twitter, and YouTube from legal liability for the vast majority of content posted by their users. The law was designed to protect fledgling internet companies from being overwhelmed by lawsuits while still giving them room to moderate harmful content. As a result, almost all online speech is permissible in the US, with some key exceptions for content related to sex trafficking, child pornography, and intellectual property violations.

The Cohen v. California ruling in 1971, which shaped much of the American free speech doctrine, encapsulates this mindset. Justice John Marshall Harlan II famously wrote: *"The constitutional right of free expression is powerful medicine in a society as diverse and populous as ours. It is designed and intended to remove governmental restraints from the arena of public discussion, putting the decision as to what views shall be voiced largely into the hands of each of us."* This reflects the bedrock principle that online platforms should not be treated as publishers responsible for the speech of others, enabling a relatively unregulated internet.

However, this laissez-faire model, while protecting free speech, has also enabled the spread of misinformation, disinformation, and malinformation (MDM), which can disproportionately harm marginalized communities. As of now, Twitter has banned political ads entirely, while Facebook has resumed allowing political ads but with restrictions on targeting sensitive attributes like race and religion. Platforms like Spotify have also recently reinstated political ads after a temporary ban. However, the information overload caused by social media creates a chaotic environment where it becomes challenging to distinguish fact from fiction.

Bad actors often exploit this situation. Marginalized communities, especially Black, Latino, low-income, and immigrant groups, are frequently targeted with disinformation aimed at voter suppression. For instance, in Alabama's 2017 US Senate special election, African American voters in Jefferson County were sent false text messages about polling site changes. Similarly, in the 2010 Maryland gubernatorial election, campaign workers used robocalls to mislead Black voters into thinking the election was already decided, urging them not to vote. These incidents highlight the real-world consequences of disinformation and raise questions about the adequacy of current laws like Section 230. Critics argue that the law provides too much protection to platforms, allowing them to evade responsibility for the disinformation or harmful content they host, particularly when it leads to real-world harm like voter suppression.

The ongoing Gonzalez v. Google case may further complicate the discussion around Section 230. This landmark case tests whether platforms are liable for amplified content, such as personalized recommendations through features like "You Might Like" or "Up Next." The plaintiffs argue that while platforms may be immune for simply hosting content, they should be held accountable for highlighting or promoting harmful content, especially if it contributes to real-world violence, as was allegedly the case with ISIS propaganda on YouTube. This case, along with similar lawsuits targeting Twitter, signals a potential shift in how courts view the scope of Section 230 immunity.

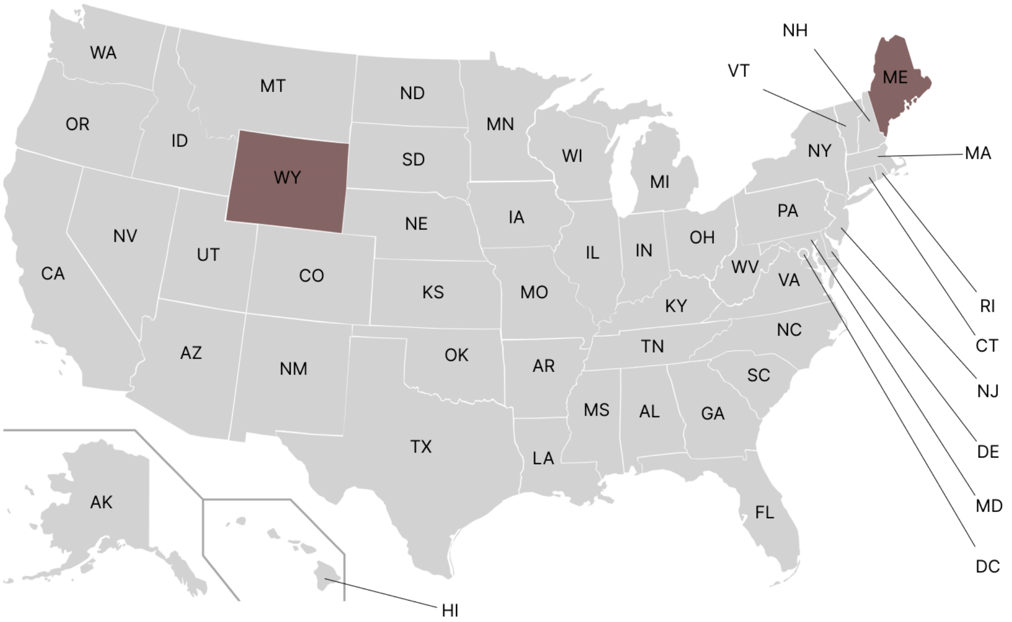
Reform Proposals: Balancing Free Speech and Accountability

In light of these ongoing controversies, several reform proposals have emerged in Congress that aim to narrow Section 230 protections without entirely dismantling the law. One such proposal is the SAFE TECH Act, which seeks to remove liability protections for paid content such as political ads, ensuring that platforms cannot profit from harmful or fraudulent advertisements. This bill also aims to hold platforms accountable for civil rights violations and wrongful death claims, areas where Section 230's immunity has been criticized for allowing platforms to escape legal responsibility.

The PACT Act takes a slightly different approach, focusing on transparency and accountability. It proposes that platforms be required to publish clear content policies and provide periodic transparency reports detailing their content moderation practices. This reform effort is intended to give users more insight into how platforms decide what to remove, allowing for more accountability. There are also efforts to address specific content categories. For example, FOSTA-SESTA passed in 2018 carved out an exception for sex trafficking, removing immunity for platforms in cases where they enable or facilitate trafficking. Other proposals target harmful content such as cyberbullying, doxxing, and online harassment, seeking to narrow Section 230's protections in these areas.

The debate over Section 230 reform is highly polarized. Democrats generally push for stricter platform liability to combat disinformation, hate speech, and other online harms, while Republicans often advocate for less moderation, claiming that platforms suppress conservative voices. This polarization has made bipartisan reform difficult despite both sides agreeing that some form of change is necessary. However, the stakes are high. Without Section 230, the internet could fundamentally change. Due to legal risks, platforms might be forced to either heavily censor content or stop hosting user-generated content altogether. Smaller platforms, in particular, might not be able to survive such a legal environment, leading to less diversity in online spaces.

While Section 230 remains a powerful federal law that shapes online content regulation, its limitations have become increasingly apparent, especially in addressing the rapid spread of disinformation, harassment, and harmful content. As debates around federal reforms continue, individual states have begun taking matters into their own hands, crafting legislation to address the gaps left by federal law. States are focusing on areas such as platform accountability, election misinformation, cyberbullying, and privacy protection, aiming to create tailored solutions that respond to the specific needs of their jurisdictions. This shift to state-level legislation shows a growing recognition that federal laws alone may not fully address online harms' complex and evolving nature. In the next section, we will explore how various states are stepping up with innovative laws and policies that seek to mitigate the shortcomings of existing federal regulations.



States with No MDM Laws

Federal Laws on MDM Regulation

The federal regulation of misinformation, disinformation, and malinformation (MDM) in the US operates within a patchwork of existing laws that intersect with various policy areas. While no comprehensive federal legislation specifically regulates MDM, several laws and initiatives indirectly address the spread of harmful content online.

1. **Public Health (COVID-19 Misinformation)**: During the COVID-19 pandemic, the federal government made significant efforts to curb misinformation, particularly concerning vaccines and public health protocols. The Biden administration worked closely with social media platforms to identify and flag harmful misinformation, while the Surgeon General actively led public campaigns to combat disinformation. However, legal restrictions on the federal government prevent direct regulation of the content itself, leaving social media platforms to establish their own policies in collaboration with federal health agencies.

2. **Election Misinformation**: Misinformation around elections, a critical concern in recent years, has drawn the attention of the Department of Homeland Security's Cybersecurity and Infrastructure Security Agency (CISA). Through partnerships with state and local governments, CISA has initiated several programs to monitor and mitigate election-related disinformation. However, while the federal government can address foreign interference in elections, regulatory efforts are restricted to collaboration and advisory roles rather than imposing direct mandates on platforms.

3. **Social Media Accountability**: Section 230 of the Communications Decency Act remains the core federal statute that governs platform liability for user-generated content. While reforms to Section 230 have been proposed, including carving out exceptions for misinformation or targeted ads, no substantive changes have yet passed, leaving the regulation of MDM mainly in the hands of the platforms themselves.

4. **National Security**: The US government addresses foreign disinformation campaigns primarily through national security efforts. The FBI's Foreign Influence Task Force is tasked with countering state-sponsored disinformation campaigns from actors like Russia and China. These efforts focus on preventing foreign influence operations that target US elections or spread misinformation designed to destabilize social or political structures.

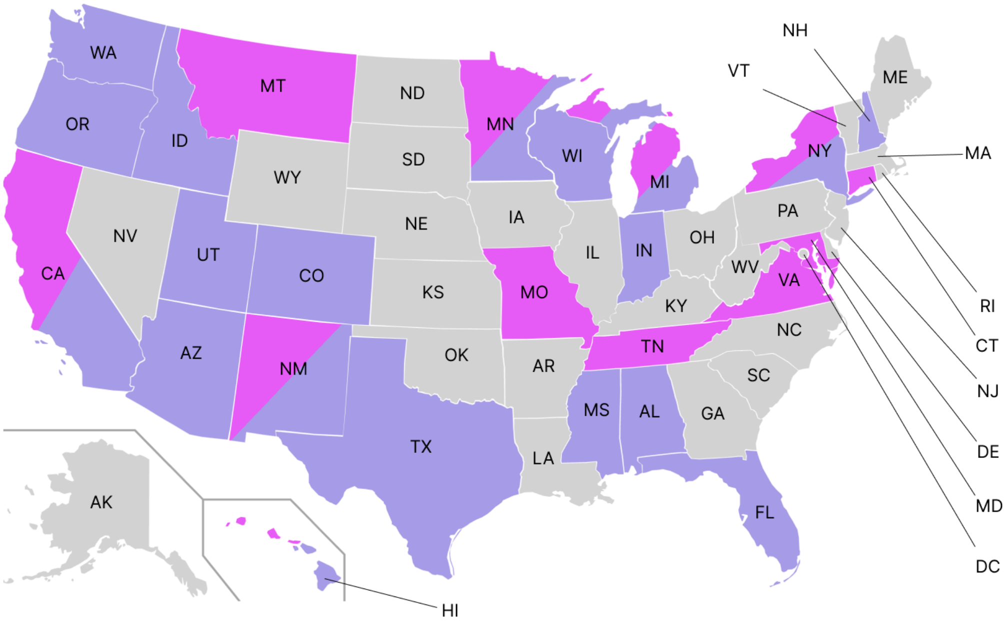
Federal MDM regulation remains a mix of advisory, collaborative, and security-related efforts with no overarching framework in place. The gaps in federal law leave much of the regulatory responsibility to states, leading to varied approaches across the country.

Key Differences Across State Regulations

State-level regulations on MDM reflect a broad range of approaches influenced by political, social, and economic factors unique to each state. Some states have taken aggressive action to tackle misinformation, particularly concerning elections and public health, while others remain cautious about introducing new regulatory frameworks, focusing instead on issues like data privacy or platform transparency. Here are some of the key differences that emerge across state MDM regulations in our research:

Focus on Privacy vs. Content Moderation: States like California and Virginia are at the forefront of privacy regulations that impact how platforms handle user data, specifically by providing them for targeted MDM campaigns. For example, the California Privacy Rights Act (CPRA) imposes strict obligations on data handling, indirectly pushing platforms to be more transparent and accountable, thereby influencing how misinformation is managed. California and New York have implemented or proposed legislation requiring platforms to publish regular transparency reports and disclose how they moderate user content. In contrast, states like Texas and Florida focus more on protecting free speech and limiting platform intervention, avoiding direct regulation of content moderation practices. These states, critical of the liberal-leaning approach of the major social media platforms, prioritize anti-censorship laws that emphasize protecting user-generated content from removal, even when it may include misinformation. The emphasis is to ensure platforms do not engage in what they perceive as politically motivated censorship.

Focus on Election vs. Public Health: States approach MDM regulation with varying priorities, focusing on areas most vulnerable to misinformation. For example, states like Minnesota, Michigan, and New Mexico have implemented laws targeting both traditional election misinformation and AI-generated content, including deepfakes, to address false claims about voting procedures and prevent voter suppression. Conversely, states like New York and Massachusetts concentrate on public health misinformation, especially around COVID-19 and vaccines, aligning these regulations with federal health initiatives to counter disinformation that could harm public health. This divergence underscores a fragmented regulatory landscape in which states independently address MDM according to their unique priorities and concerns.



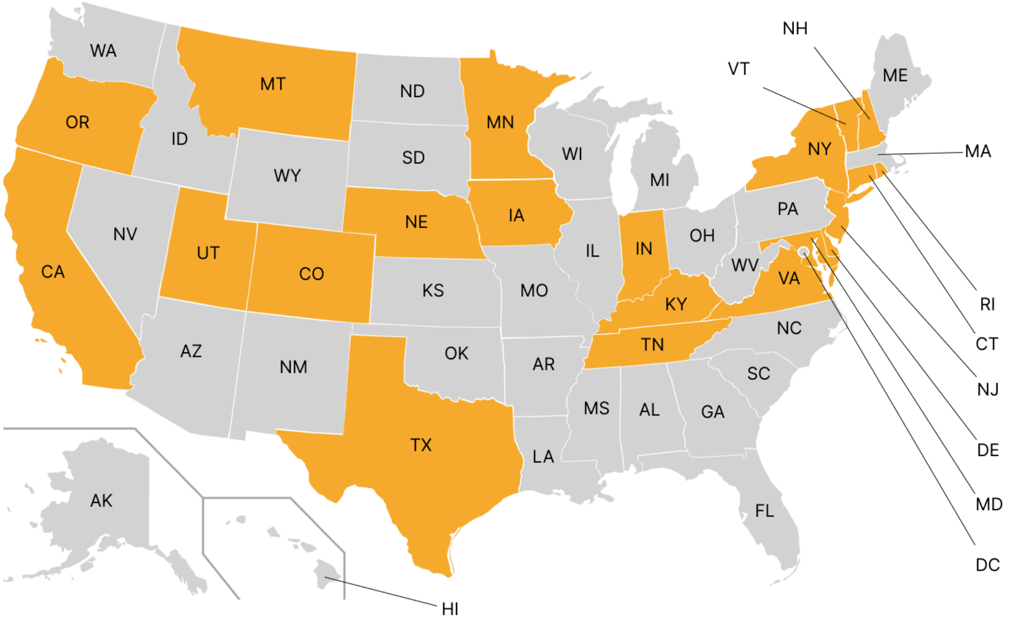
Election MDM Laws (AI: Purple and Non-AI: Pink)

Thematic Breakdown of US Legislation Addressing Misinformation and Disinformation

It is essential to categorize these laws thematically to address the complex landscape of laws regulating misinformation and disinformation (MDM) in the United States. The diversity in state and federal regulations is significant, reflecting varying priorities, from protecting individual privacy to safeguarding elections or national security. For clarity and analysis, we have grouped relevant legislation into seven categories that each play a unique role in addressing the spread and impact of MDM. These categories explore the multidimensional approach required to combat MDM while balancing free speech and civil liberties.

1. Privacy and Data Protection

Privacy and data protection laws are crucial in regulating the flow and use of personal data, which often becomes a key vehicle for spreading misinformation and disinformation. These laws limit how companies, platforms, and advertisers can collect, store, and exploit user data, thus indirectly curbing the spread of harmful content. The California Consumer Privacy Act (CCPA) and California Privacy Rights Act (CPRA) are significant because they grant consumers more control over their personal information, forcing platforms to be more transparent about data practices. These laws allow individuals to know what data is being collected and give them the right to have that data deleted. This transparency makes it more difficult for bad actors to use targeted data-driven campaigns to spread disinformation.



Privacy and Data Protection Laws

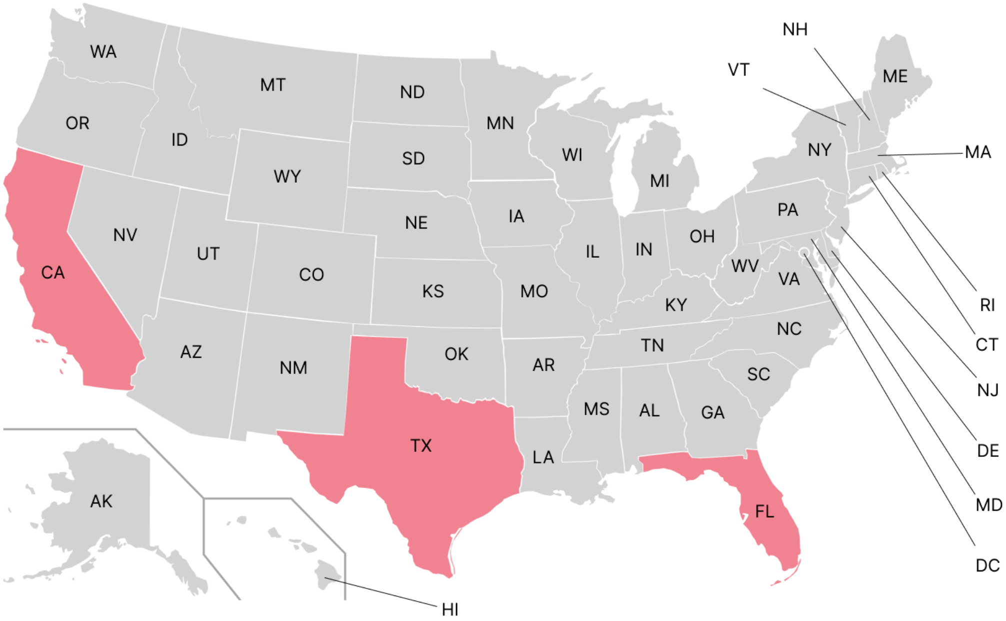
Other states, such as Virginia, Colorado, and Utah, have adopted similar frameworks that give consumers rights over their personal data and impose obligations on businesses to protect that data. These laws aim to limit unauthorized access and prevent the use of personal data in spreading false narratives or malicious misinformation campaigns. At the federal level, sector-specific privacy regulations like HIPAA (for health information) and GLBA (for financial data) ensure that sensitive personal data is protected. While these laws are not directly focused on MDM, they help limit the use of personal information in campaigns targeting individuals based on sensitive or personal data.

Privacy-focused applications can prevent individuals from being targeted by MDM campaigns. Applications emphasizing absolute privacy and the prevention of data leaks ensure that users' personal data is shielded from third-party surveillance. However, privacy cannot be effectively addressed on an individual level alone. Even if one person's data is fully protected, MDM campaigns can still target them indirectly by exploiting the data of others within their network who share similar traits, such as location, interests, or identity categories. Increasing the number of people using privacy tools strengthens collective privacy, making it harder for disinformation campaigns to leverage aggregated data for targeted manipulation.

2. Transparency, Platform Accountability, and Anti-Censorship

US states have adopted varying approaches to regulating transparency and accountability on social media platforms, especially about misinformation, disinformation, and content moderation practices. California's AB 587, effective January 2024, is a leading example of transparency-focused legislation. Social media companies generating over $100 million in revenue must publicly disclose their terms of service and submit biannual reports to the California Attorney General. These reports will detail how the companies manage hate speech, extremism, disinformation, and other harmful content, ensuring platforms are more accountable in their moderation practices. While initially aimed at data collection, AB 587 is expected to lead to stricter regulations over time.

In contrast, states like Texas and Florida have prioritized anti-censorship laws. Texas HB 20 and Florida SB 7072 prohibit social media platforms from censoring content based on political viewpoints. These laws allow users and political candidates to sue platforms if they believe they've been wrongfully censored. However, both laws have faced challenges on First Amendment grounds and are currently under review by the US Supreme Court. The contrast between California's transparency-driven regulations and Texas and Florida's anti-censorship laws reflects the broader debate over the role of social media in public discourse. As these state laws continue to evolve, they highlight the complexity of balancing free speech with the need for accountability in combating harmful content online.

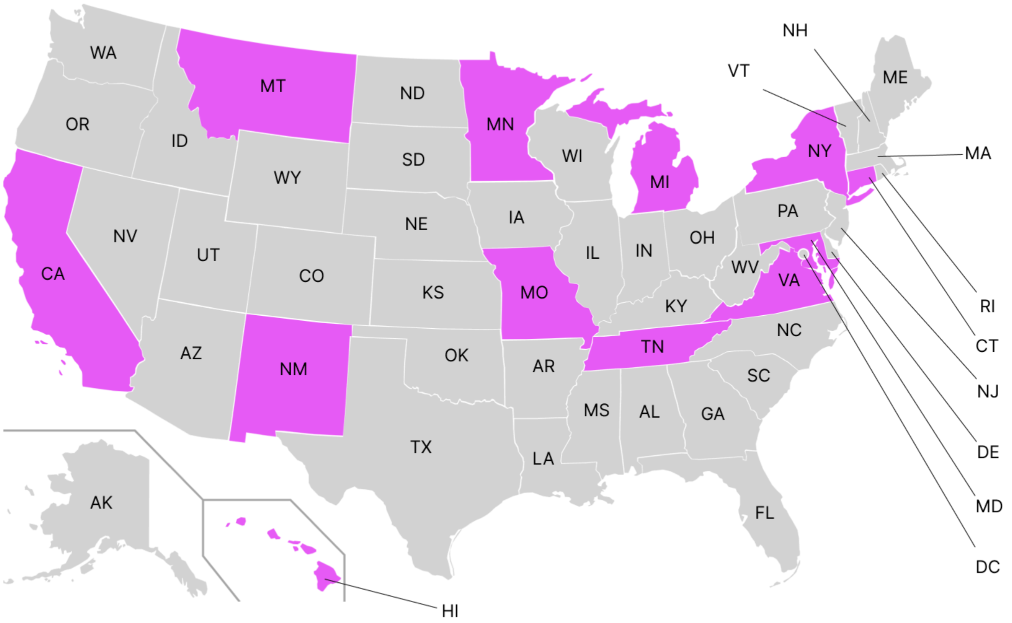


Transparency, Platform Accountability, and Anti-Censorship Laws

3. Election Misinformation (Excluding AI)

Election misinformation has become a significant concern in recent years, particularly surrounding the 2016 US presidential election, where millions of instances of disinformation were recorded, especially regarding vote-by-mail policies. This spread of election-related falsehoods has prompted states to take legal measures to protect the integrity of the electoral process. This section does not cover AI-generated election misinformation, which will be addressed separately in the following category.

Various states have enacted laws to combat election misinformation by targeting false claims about voter eligibility, voting processes, and election outcomes. For example, Connecticut's election law prohibits misrepresenting absentee ballot eligibility to potential voters (Conn. Gen. Stat. § 9-135). In Hawaii, a recently enacted law (Haw. Rev. Stat. § 11-A) prohibits the reckless distribution of materially deceptive media intended to harm a candidate's reputation or alter voting behavior during the election period, with penalties including misdemeanors or felonies depending on the severity of the violation. Similarly, Maryland (Election Law § 16-101) and Michigan (MCL - Section 168.932f) have laws that address falsifying voter registration information or distributing deceptive media about candidates within specific timeframes before elections.



Election MDM (Excluding AI) Laws

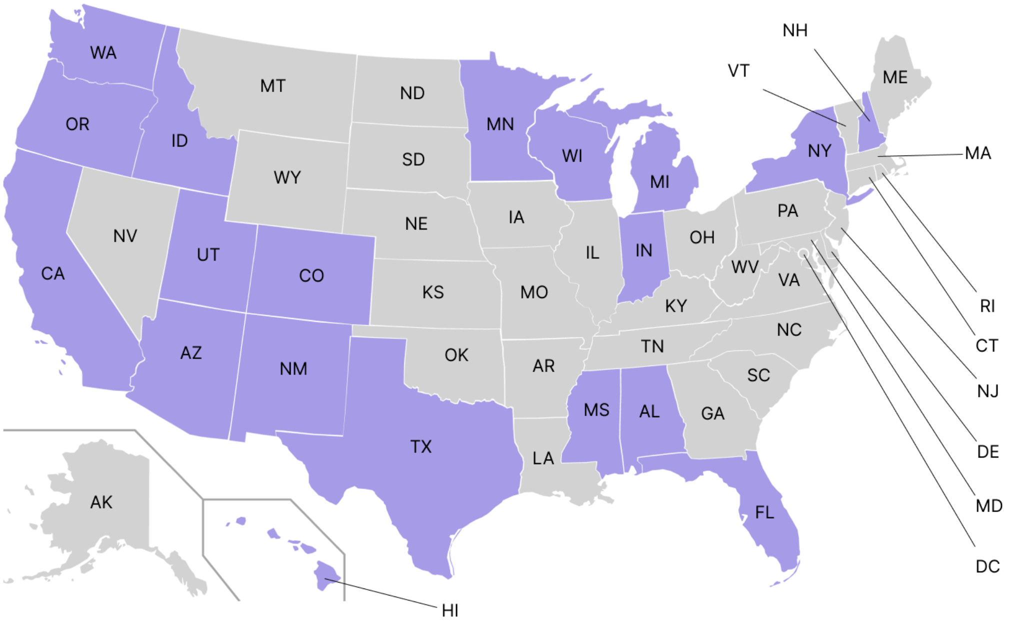
Several states, such as Tennessee and Montana, have also introduced statutes criminalizing the spread of false information about voting procedures, including misleading voters about voting times, places, and eligibility requirements. These laws aim to prevent voter suppression through disinformation that could discourage or mislead voters about their right to participate in elections.

4. AI (Election-Specific) Laws

The rapid development of artificial intelligence (AI) has introduced significant challenges in regulating election-related content, especially in combating AI-generated disinformation, commonly referred to as “deepfakes.” These AI-generated images, audio, or videos can create highly realistic yet entirely fabricated depictions of candidates, posing a severe threat to the integrity of democratic processes. Legislators across the US have begun to address these urgent threats with targeted laws, focusing on preventing AI-manipulated media from influencing election outcomes.

Several states have taken proactive measures by passing or proposing laws that focus specifically on deepfakes in the context of elections. California's laws, for example, require labeling AI-generated political content and imposing penalties for distributing deepfake media aimed at misleading voters about a candidate's actions or statements during the election period. Similar laws have been enacted in Florida, Minnesota, and Texas, requiring political deepfakes to be clearly labeled or banned outright in political campaigns. States like Arizona have gone further by creating civil remedies for individuals harmed by AI-generated misinformation. These measures should reassure the public that steps are being taken to protect the integrity of elections.

In addition to state actions, federal lawmakers have also recognized the growing threat posed by AI-generated election content. The bipartisan AI Labeling Act of 2023 was introduced to require clear and conspicuous labels on generative AI content, including political communications, to ensure transparency in election-related media. While this bill remains in committee, it underscores the broader legislative push to tackle AI-driven election disinformation.



AI (Election-Specific) Laws

One of the main challenges in regulating AI-generated election content is ensuring that laws are both effective and constitutional. Some laws mandate disclaimers on synthetic media created within a specific period before an election, while others allow for penalties based on the intention behind the content's distribution. For instance, laws in

Mississippi and Illinois criminalize the use of AI to spread false information intended to manipulate voter behavior or harm candidates' reputations.

As the 2024 elections approach, state and federal governments closely watch how AI-generated disinformation might impact voter decisions. These legislative efforts aim to balance protecting free speech and ensuring that voters are not deceived by fabricated content. The evolving legal landscape around AI in elections will likely continue to shape how AI technologies are regulated to safeguard democratic processes.

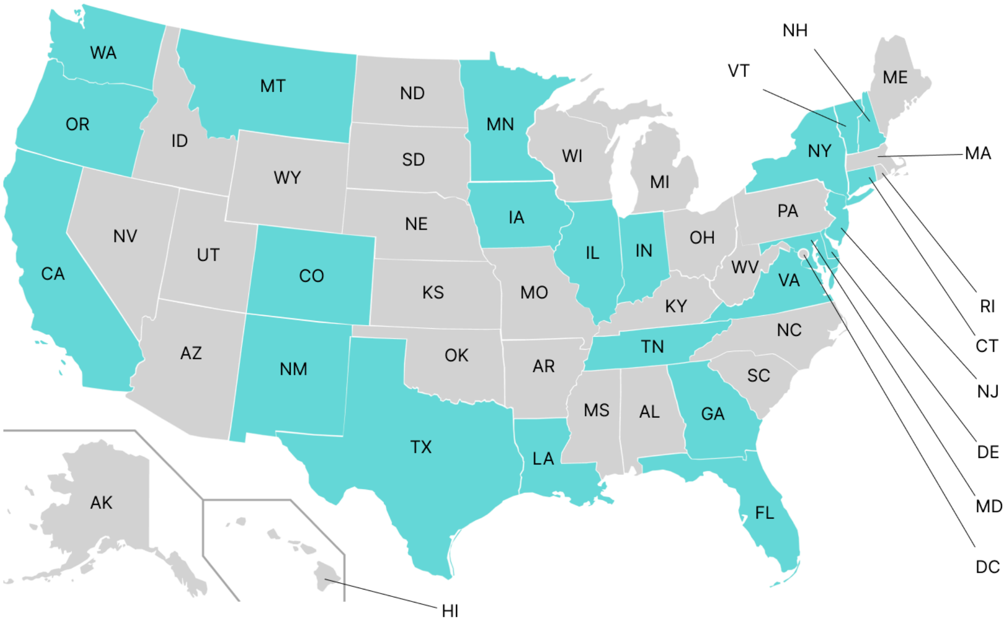
5. AI Regulations (Excluding Elections)

While no comprehensive federal legislation directly regulates artificial intelligence (AI) in the US, several federal initiatives have emerged to provide guidelines and shape policy. At the federal level, the SAFE Innovation AI Framework outlines non-binding principles for AI safety and innovation, while the REAL Political Advertisements Act aims to ensure transparency in AI-generated political content. Other legislative proposals like the Stop Spying Bosses Act and the Draft No FAKES Act focus on preventing misuse of AI in the workplace and protecting individuals' likenesses from unauthorized AI-generated recreations.

The AI Research Innovation and Accountability Act is another federal proposal to enhance transparency and accountability for high-risk AI systems by establishing enforceable testing and reporting standards. Meanwhile, the Blueprint for an AI Bill of Rights and President Biden's Executive Order on Generative AI (2023) push for responsible AI development, emphasizing fairness, transparency, and privacy. Regulatory bodies such as the Federal Trade Commission (FTC) and the National Institute of Standards and Technology (NIST) have also taken steps to apply existing consumer protection and risk management standards to AI technologies. Notably, the Federal Communications Commission (FCC) expanded the Telephone Consumer Protection Act (TCPA) in 2024 to regulate AI-generated voices in robocalls, further demonstrating the federal government's gradual approach to addressing AI use.

The lack of a federal AI law has prompted many US states to take the lead in regulating AI, with each state adopting unique approaches focusing on specific sectors. States like California, Colorado, New York, and Texas have become pioneers in AI regulation, often targeting privacy, transparency, and discrimination issues. California, the world's fifth-largest economy and home to many leading AI companies, has introduced several

AI-related bills. One such effort, the Artificial Intelligence Accountability Act, mandates that state agencies produce reports assessing the risks of generative AI. California has also proposed laws like SB 892 requiring AI service providers to enter public contracts with the state to meet privacy, safety, and non-discrimination standards. The state is also pushing for establishing the California Artificial Intelligence Research Hub; a collaborative entity focused on ensuring that AI development in the public sector balances innovation with privacy and security concerns. These initiatives are part of California's broader strategy to regulate AI while fostering responsible innovation.



AI (Excluding Elections) Laws

New York is similarly active in regulating AI, particularly in the context of automated decision-making. Its Automated Employment Decision Tools (AEDT) Law places strict transparency requirements on employers who use AI to evaluate job candidates, ensuring that AI systems are unbiased and fair. Additionally, New York has modeled new AI regulations on Colorado's insurance laws, restricting insurers' use of AI-driven data like credit scores and social media habits to prevent discrimination. Colorado has led with its SB-169, a law prohibiting insurers from using biased data collected by AI to determine insurance premiums. This law reflects growing concerns over algorithmic discrimination in the insurance industry and has become a model for similar legislation in other states.

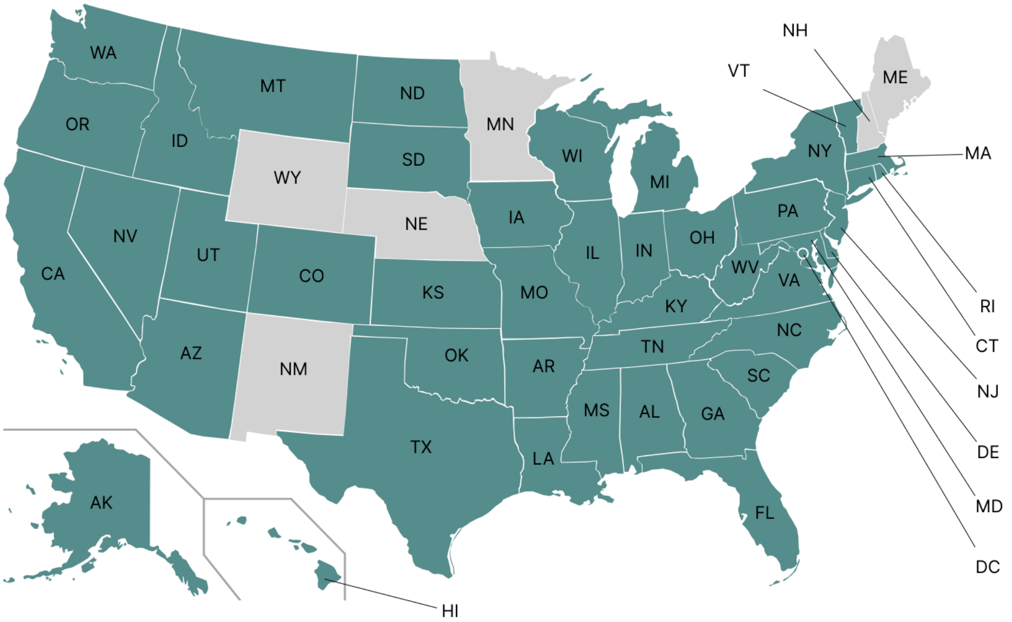
Texas has implemented HB 2060, which established an AI advisory council to study the impact of AI technologies on state agencies and public services. The council is tasked with issuing recommendations to ensure that AI systems used by the government are transparent, fair, and protective of individual rights. Florida has vehemently opposed AI transparency, particularly in the public sector. Proposed laws like the AI Transparency in Government Technology Act call for the creation of a council to monitor AI deployment across government agencies and issue regular reports on how these technologies impact privacy and civil liberties.

Despite these varied efforts, many states share common themes in their regulatory approaches. States like Illinois, Virginia, and Washington have introduced laws informing individuals about how AI is used in decision-making processes, particularly in employment, insurance, and government services. These states have also focused on data privacy protections, often giving consumers the right to opt out of AI-driven profiling or automated decision-making. While the US lacks a comprehensive federal AI law, state-level initiatives are shaping the landscape by targeting specific AI applications in the employment, insurance, and public services sectors. This patchwork of regulations reflects a growing consensus that AI technologies must be governed by principles of transparency, accountability, and fairness, with states like California, New York, Colorado, and Texas leading the charge.

6. Cyberbullying, Defamation, and Online Harassment

In the United States, nearly all states have enacted laws to address cyberbullying and online harassment. While these laws, which often target ongoing harassment or cyberstalking, vary in their definitions and penalties, they are essential for combating the rise in electronic abuse. At the federal level, the Stalking and Harassment Laws (18 U.S.C. § 2261A) make it a crime to harass or intimidate someone using electronic communication across state lines, directly addressing online harassment and cyberstalking. However, despite the national push for regulation, six states — Maine, Minnesota, Nebraska, New Hampshire, New Mexico, and Wyoming — do not currently have specific laws criminalizing cyberbullying. In these states, victims often rely on more general anti-bullying laws, school policies, or defamation laws to pursue justice.

The tragic suicide stories of individuals like Megan Meier and Tyler Clementi have driven much of the legislative momentum, leading many states to enact stricter cyberbullying laws in the wake of high-profile cases. While some states have taken steps to introduce severe penalties, including criminal charges like involuntary manslaughter in extreme instances, the patchwork of state laws still leads to inconsistencies across the nation. Some states, like Texas, have created broader statutes like “David’s Law,” which includes civil remedies and amended criminal harassment codes to address electronic communications.



Cyberbullying, Defamation, and Harassment Laws

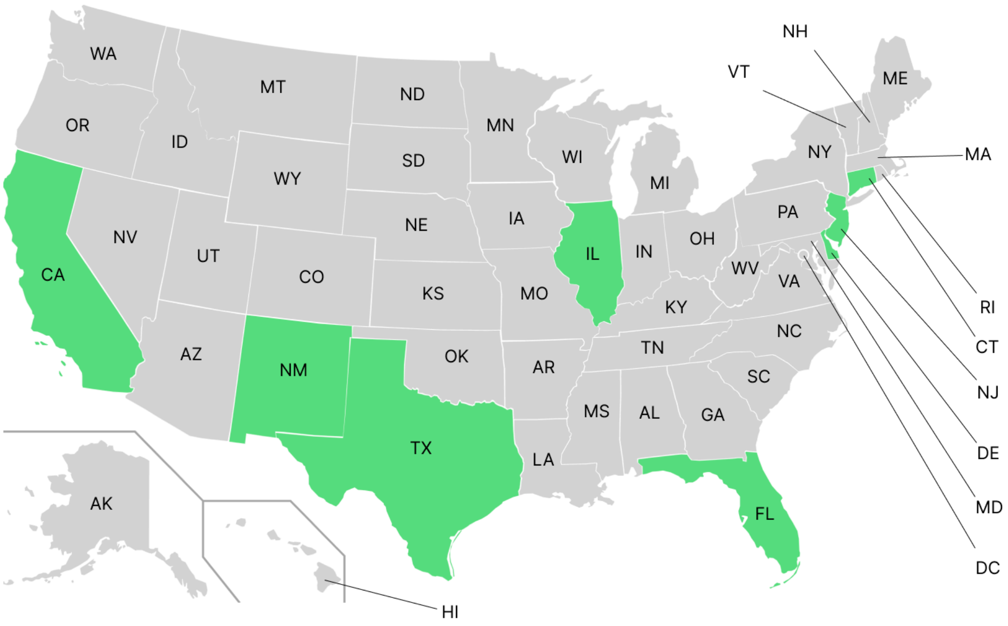
Although federal legislation directly criminalizing cyberbullying has not yet been enacted, the continued advocacy for a national standard, driven by the need to protect vulnerable individuals from online abuse, points toward possible future reforms. For now, states bear most of the responsibility, with each approaching the issue through varied legal frameworks that reflect differing levels of strictness in how they penalize cyberbullying.

7. Digital Literacy and Public Education

Digital Literacy and Public Education have become focal points for many states as they grapple with the rise of misinformation and the challenges of the digital age. Although no federal law explicitly mandates media literacy education, several states have taken the initiative to address this growing concern, tailoring their approach to local needs.

States like New Jersey, Delaware, Illinois, Texas, and California have passed laws emphasizing the importance of teaching media literacy in public schools. New Jersey, a frontrunner in this area, passed a law in 2023 that requires K-12 students to receive instruction on media literacy, focusing on distinguishing facts from opinions and evaluating the credibility of information. Similarly, Delaware's Digital Citizenship Education Act mandates that all schools implement media literacy standards. In Illinois, high school students are now required to take a media literacy unit that examines the impact of media on society. At the same time, Texas has introduced a law focusing on misinformation and social media's role in shaping public opinion. California recently passed a comprehensive law integrating media literacy into curriculum development and teacher training to foster critical thinking skills in K-12 students.

As online misinformation continues to increase, particularly targeting younger audiences, these collaborative efforts are crucial in preparing students to navigate the complexities of the digital world. Educators and lawmakers recognize the importance of teaching students how to evaluate the information they encounter critically, equipping them with tools to recognize bias, credibility, and manipulation.



Digital Literacy and Public Education Laws

Privacy: The Key to Combating Misinformation in a Digital Age

Addressing the challenges of misinformation, disinformation, and malinformation (MDM) requires a multidimensional approach. Legal reforms can strengthen existing protections against deceptive practices, especially in the context of elections, and empower election officials to act swiftly in the face of disinformation. Similarly, internet platforms must improve content moderation practices while maintaining transparency, and public education initiatives should continue to build media literacy to help individuals recognize and resist harmful content. These steps are crucial but should not be the sole focus in managing MDM.

At the core of a robust defense against MDM is the empowering role of privacy. Privacy plays a critical role in limiting the effectiveness of disinformation campaigns by protecting personal data, which bad actors exploit to tailor and target harmful narratives. The more individuals secure their privacy, the harder it becomes for disinformation agents to profile and manipulate them. This empowerment through privacy creates a decentralized defense mechanism that does not depend on government overreach, reducing the risk of censorship under the guise of combating MDM.

Furthermore, privacy isn't just an individual concern; it's a collective matter. When widely adopted, privacy tools create a collective barrier, preventing MDM campaigns from precisely targeting individuals based on shared traits, locations, or interests. This collective approach, spearheaded by privacy tools, strengthens our digital ecosystem and makes everyone part of a larger protective network, shielding us from data-driven disinformation efforts.

Ultimately, privacy provides a balanced solution that safeguards individuals and the integrity of information without the need for excessive state intervention. In a world where disinformation is becoming increasingly sophisticated, privacy is not just a shield — it's an essential part of the fight for a trustworthy and transparent flow of information.

Download the full dataset here for a comprehensive database detailing laws in all 50 states in each of the seven MDM categories.