

Big Tech Won't Fix Disinformation Without Regulation

AUSTIN AHLMAN
REPORTER
OPEN MARKETS INSTITUTE

COURTNEY RADSCH, PhD
DIRECTOR
CENTER FOR JOURNALISM & LIBERTY

JUNE 2024



OPEN MARKETS
LIBERTY * DEMOCRACY * PROSPERITY

The sheer volume of misinformation proliferating on major platforms controlled by dominant Big Tech companies like Google and Meta is alarming. Also of profound concern is the inconsistency and lack of transparency in how content is served and moderated, contributing to the steep erosion of social trust around the world. The glut of low-quality information, whether created with the intention of shaping public discourse or for monetary gain, is in large part driving the [decline in trust](#) we have in our institutions and exacerbating the political polarization occurring in most democracies.

A growing body of evidence suggests that U.S. geopolitical rivals like Russia and China are intentionally running large-scale information operations on the major platforms in order to hobble public discourse and democratic conflict resolution. That said, state-sponsored information activities are not limited to authoritarian rivals, with evidence showing a rapidly growing number of nations turning to platform-based propaganda and information manipulation as they seek to influence public opinion and pursue various domestic and geopolitical agendas. (It is worth noting that the United States and its democratic allies [also participate](#) in these kinds of campaigns and are often subject to less scrutiny as the host nation of most of the major platform companies.)

Along with coordinated state-aligned campaigns, the business models of search and social media platforms often encourage and reward disinformation while undermining high-quality and public interest information, such as journalism. Financial incentives that reward engagement regardless of the quality of information or its origin have given rise to a global industry trafficking in low-quality information. These incentives continue to enable large and growing networks of content farms, which create and disseminate false or misleading information not to persuade but to profit from the advertising revenue generated by their posts. As an explosive [report](#) from the MIT Technology review demonstrated, content farmers in nations like Moldova and Kosovo have used this model to make millions of dollars from misinformation directly tied to democratic elections.

Despite making lofty promises, the major tech platforms' actions following the United States' 2016 and 2020 elections—which brought widespread attention to the dangers of information manipulation and the degradation of the information environment—have been inconsistent, incoherent, and inadequate. Nor have policymakers effectively reigned in the manipulative practices or problematic business models responsible for these threats.

The same disinformation tactics that malicious and profit-driven actors have utilized in recent election cycles appear to be once again tarnishing the democratic process of several nations holding elections this year. To make matters worse, several major platforms are actively suppressing or outright blocking access to high-quality information in a bid to protect their profits from new laws seeking to give journalistic organizations fair compensation for the use of content like original reporting. This follows the monopolization of digital advertising markets by Google, Meta, and more recently Amazon that has driven the destruction of the business model that fostered the production and distribution of high-quality information in decades past. The platforms' yearslong failure to address declining information integrity should be a wake-up call to regulators, who have thus far demurred from addressing discrimination against high-quality content, conflicting labeling standards, and opaque data transparency practices that have allowed disinformation to flourish.

Most of these alarming practices have been exacerbated by the mainstreaming of new generated artificial intelligence products and services, which enable increasingly convincing disinformation to be produced at a lower cost and greater scale than ever before. This dynamic has greatly expanded the capacity of disinformation operatives and content farmers alike, reducing costs for the former and increasing returns for the latter. As one particularly alarming [report](#) from Europol found last year, as much as 90 percent of the internet could be comprised of AI-generated information of questionable origins and veracity by 2026 without policy interventions. AI innovation is far outpacing government regulation, and the platforms—all of which

have substantial investments in the AI boom—have shown little appetite for meaningful self-regulation while spending outsized amounts to shape the policy debate in their favor, allowing them to maintain their dominant positions.

Fair Treatment of High-Quality Information

The primary way platforms have obstructed the availability and spread of high-quality information is through their capture of the advertising revenue streams which had long provided the foundation of most journalistic organizations. Tens of billions in annual advertising revenue that should be going to support the cost of producing journalism is instead being siphoned into the vaults of Big Tech companies through their control over and manipulation of the digital advertising market.

The Department of Justice has brought a historic antitrust case against Google over such abuses. Although the suit has been dubbed the “biggest tech monopoly trial of the 21st century,” the behavior being prosecuted is rarely given serious weight in discussions about disinformation on the platforms. But Open Markets Institute has been warning about the threats monopolization poses to the information ecosystem for years, culminating in a release of our [definitive report](#), “Democracy, Journalism, and Monopoly,” at the end of last year.

While progress on other fronts of the disinformation war has been slow, democratic governments around the world have taken the first steps toward forcing the major platforms to address the gaping holes they have created in journalistic revenue streams by negotiating fair compensation with publishers for the value their content adds to the platform. Australia kicked off this wave of reforms by implementing news media bargaining code in 2021. While challenges remain, the reform has helped stabilize the nation’s news media industry and protect hundreds of jobs for journalists producing high-quality information.

One such framework for the United States, Senator Amy Klobuchar’s Journalism Competition and Preservation Act, has drawn strong bipartisan support but has not yet become law. Meanwhile, several states are pursuing similar bipartisan legislation. As evidenced in the Center for Journalism and Liberty’s [global tracker](#), these market-based reforms have found traction around the world, as more governments recognize the need to preserve the business model for high-quality information.

But rather than deal in good faith with these efforts to address destructive imbalances, platform monopolies have escalated their attacks on high-quality information producers and demonstrated little if any concern over the effect that putting their thumb on the algorithmic scale may have on public discourse. Google and Meta have curbed or entirely cut off access to genuine news in several jurisdictions in retaliation for efforts to make sure publishers are fairly compensated. And under the direction of new owner Elon Musk, the app formerly known as Twitter has also “de-boosted” posts linking to high-quality information sources and even throttled the speed of connections to news sites he views unfavorably.

The platforms have deployed several talking points to support their attacks on news publishers and the fair compensation frameworks publishers have proposed. One is that there is no precedent for mandating payments for retransmission of content. A second is that it is impossible to negotiate with hundreds of media organizations simultaneously.

But neither of these arguments holds up to close scrutiny. In fact, one potential model for news publishers is a long-time policy that has governed the relationship between local over-the-air (or OTA) television channels and cable providers. Thanks to the Federal Communications Act’s focus on ensuring a level playing field between media organizations and distributors, cable networks like Comcast must carry all the local OTA channels in the regions in which they operate.

The Communications Act, which was originally passed in 1934, was amended multiple times in the 1980s and 1990s to account for the power of cable providers, culminating in the Cable Television Consumer Protection Act of 1992. That amendment created the [present requirement](#) for cable networks to negotiate in “good faith” with local broadcast stations for the redistribution of their content. The laws also make it clear that the Federal Communications Commission is fully empowered to step in and enforce the requirements. Though these broadcasters’ channels are often free to access over the air, the framework acknowledges that the content they provide adds substantial value to cable networks that must be shared directly with local stations. Cable has long been the primary distributor of broadcast news.

Crucially, these local broadcast stations are also entitled to prime channel slots in recognition of the role local media plays in connecting communities and fortifying our democracy. A similar model for the platforms could not only prohibit discrimination against news publishers and other producers of high-quality information but actually force platforms to give preferential treatment to high-quality information—a flat reversal of the current dynamic.

Account and Content Labeling Standards

While fixing the underlying market structures that are creating the current information crisis should be regulators’ primary concern, enhanced and standardized approaches to labeling and treating content and various news media and state-sponsored accounts are another key component in the fight against disinformation.

In the last five years, each of the major platforms began labeling some state-affiliated accounts and in some cases restrict those accounts from algorithmic amplification, monetization, or advertising. But the implementation of these policies has been highly inconsistent within and across the platforms, a dynamic that was propelled to the front pages with the Russian invasion of Ukraine. Each of

the platforms has different definitions of state-affiliated or state-sponsored accounts and how those accounts are treated on their service. This practice is both inadequate and susceptible to politicization, as when Elon Musk [labeled](#) some of the most respected publicly financed media outlets in the world—NPR and BBC—as state media. Meanwhile, unlabeled state-affiliated accounts can earn revenue and pay for promotion of their content, meaning that in some cases their propaganda efforts are effectively being subsidized by the platforms themselves.

There is wide variability in how the platforms currently treat state-aligned accounts, verified news organizations and other high-quality information sources, as well as how they handle AI-generated content. While most official governmental agencies appear to be labeled as such, their media outlets often are not, and the criteria platforms use to determine whether an account is state-linked are ambiguous. Several, but by no means all, state-affiliated media outlets have received the label, while media in state capitalist systems, where corporations are de facto arms of the state, have not received labels.

Apart from several states under U.S. sanctions, states or governments are allowed to operate accounts on U.S.-based social media platforms even if those same platforms are censored in their own countries. Consistent labeling and transparency are also key in light of mounting evidence that the platforms have failed to stop information operations from utilizing paid advertising programs to amplify their content and precisely target their audience. While the platforms promised reforms to advertising oversight following the 2016 elections—specifically by labeling and tightly monitoring ads containing political content—Russia has continued to utilize Facebook’s ad marketplace to interfere in public discourse around the globe, including in the run-up to the recent European Parliament elections, according to a [report](#) from last month from watchdog groups AI Forensics and CheckFirst. According to researchers, Russian operatives utilized hundreds of innocuous accounts to sponsor targeted advertisements pushing its preferred narrative about topics like Ukraine’s potential membership in the European Union and member

nations' ongoing financial support for the beleaguered nation in the face of Russia's ongoing assault. While the actual electoral impact is unclear, researchers determined the illegitimate ads reached millions of Europeans in the critical weeks leading up to the election.

Platforms have also historically used verification badges to help users understand when content is being shared by sources whose identities were verified, like journalists, news organizations, or politicians. Verified accounts received preferential treatment on many platforms, ranking higher in algorithmic amplification and recommender systems, for example. While the labels were awarded inconsistently, they were awarded to low-information content spreaders far less frequently prior to 2022.

Elon Musk's takeover of Twitter scrambled this calculus after it made verified status a pay-for-play scheme that had no bearing on the veracity of the account—such as when verification for an account falsely claiming to be Massachusetts Senator Ed Markey was approved in late 2022—and triggered a race to the bottom as Facebook and Instagram also monetized verification, turning it into another revenue source rather than a public service. The increasingly mixed signals sent by verification markers has increased the reach and apparent credibility of disinformation spreaders, while simultaneously burying high-quality information shared by high-quality sources who refuse to pay more to the very platforms that have gutted their revenue models.

The AI boom—which eased the generation of propaganda and deepfakes that have been used to mislead, impersonate, and harass—prompted new labeling practices meant to distinguish content created by humans from content created using artificial intelligence tools. Some platforms allow you to voluntarily disclose whether content has been manipulated, a policy unlikely to affect either the propaganda or economic incentives behind the use of AI-generated content.

Facebook announced efforts earlier this year to automatically mark images as AI-manipulated if its

metadata indicated as much. TikTok became the first platform to announce it was taking a similar approach with video content earlier this year. However, these approaches do not address economic incentives and rely on the voluntary adoption and inconsistent implementation of these standards by AI companies. Nor do the platforms require labeling or prohibit the removal of watermarks or other forensic identification of AI manipulation in their terms of service. Propagandists and content farmers can easily bypass these by using tools from corporations that have not adopted the voluntary standards or by creating new tools that remove this embedded data, a practice that currently carries no legal or platform-enforced penalties.

Other tech giants appear more reluctant to embrace AI disclosures. For instance, Google has been mostly silent on the matter, and the search giant has yet to announce any meaningful labeling strategy for content made outside of its own proprietary tools, despite its image and web search results becoming increasingly littered with low-quality AI-generated content.

But American users do not have to settle for a patchwork labeling regime. Instead, regulators can set standards like prominent disclosure requirements, mandating industry-wide adoption of labels, watermarks or other embedded data practices that facilitate the identification of AI-generated content, and strict penalties for those who remove that identifying data or create tools that seek to circumvent embedding requirements.

Several bipartisan legislative efforts have been introduced to penalize the creation and distribution of AI deepfakes, especially those that are sexually explicit. And with the introduction of the AI Labeling Act last year, Hawaii Senator Brian Schatz has taken the lead in crafting a broader system to label all content created or manipulated with AI tools. His bill would instruct the Federal Trade Commission to create and enforce label policies as part of its mandate to regulate unfair and deceptive business practices—a power the agency is already using to investigate many of the companies creating these tools. Presumably, enough evidence and enforcement actions

may eventually empower the FTC to pursue labeling rules through existing rulemaking authorities, as they have in many other instances where emerging technologies require fresh looks at pre-existing laws.

Perhaps more fundamentally, given the existing evidence of safety concerns and consumer harms that are already present, policymakers should consider outlawing certain types of products and services without appropriate safeguards or risk mitigation. For voice cloning and apps that produce nude likenesses, for example, there may be very limited or no legitimate use cases.

Enhanced Algorithm and Data Transparency

Other foundational reforms that would aid regulators' information integrity efforts include the mandating of greater algorithmic transparency and data access. Such reforms would enable researchers to study how the content users interact with is moderated by the platform. It would also allow them to better study how disinformation spreads, the role of platform practices and business models in perpetuating its spread, and which groups have been targeted by which actors.

Algorithmic transparency is also key to understanding how different categories of accounts and content are treated by platforms. Whether state-linked accounts are treated similarly to independent accounts is one particularly important question. None of the platforms have publicly addressed this question, likely because most moderation of state-linked accounts appears to be reactive (i.e. restricting Russian state-affiliated accounts after the invasion of Ukraine) rather than proactive or rules-based.

Mandating the creation and maintenance of comprehensive ad libraries, which enable users and researchers to view searchable databases of certain types of sponsored content, is another key transparency initiative that could be expanded and standardized. Currently, only Google and Meta have ad libraries available to researchers, and those libraries only contain small subsets of ads, such as those identified as political. Other major platforms like Twitter and Amazon have either declined

to create libraries or have provided only the most rudimentary of tools to research the ad campaigns that they have hosted, which, as previously mentioned, are increasingly being used by state-aligned accounts to spread propaganda.