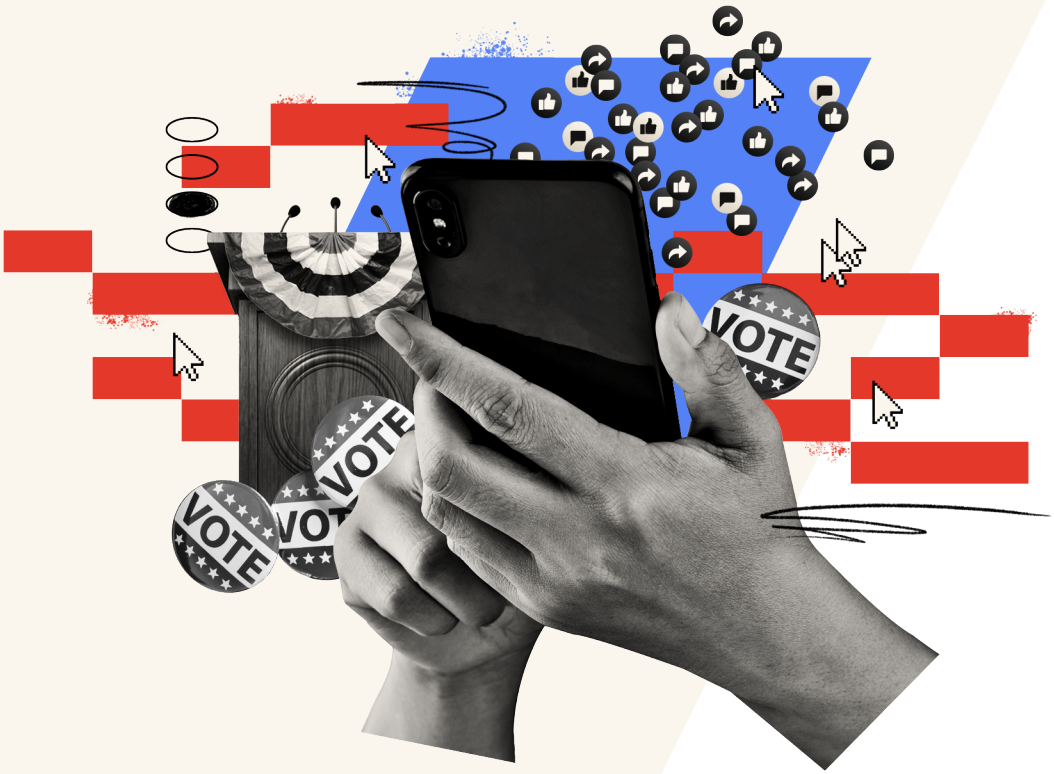
The Shortlist: Seven Ways Platforms Can Prepare for the 2024 U.S. Election

An ounce of election risk mitigation is worth a pound of cure.

MARCH 2024



Protect Democracy

Authored by: Nicole Schneidman

With contributions and thanks to: Aaron Baird, Alexandra Chandler, Chris Crawford, Rachel Goodman, Brad Jacobson, Christian Johnson, Jess Marsden, and Cecelia Vieira

Protect Democracy is deeply grateful for the expertise generously provided by the Trust & Safety professionals, technologists, election experts and scholars whose work and reviews helped to shape this report. Notwithstanding their generous input, Protect Democracy takes sole responsibility for the content of this report.

Suggested citation: Protect Democracy, *The Shortlist: Seven Ways Platforms Can Prepare for the 2024 U.S. Election* (March 2024)

This publication is available online at: protectdemocracy.org/work/platforms-prepare-2024-election

Please direct inquiries to: press@protectdemocracy.org

Introduction

The consequential choices facing tech platforms in the "year of elections"

In 2024, the "year of elections,"[1] the technology platforms that comprise today's online information ecosystem are facing a watershed moment. With the U.S. election season underway and 82 other elections being held around the world this year, platforms' election preparations and guardrails are poised to play a critical role in the production and spread of election information for more than four billion voters.[2]

Protect Democracy has produced four recommendations for each of three platform categories, (1) social media platforms, (2) messaging platforms, and (3) generative AI platforms, to inform their preparations to safeguard the information ecosystem surrounding the U.S. general election. These recommendations are not intended to be comprehensive; rather, they are priority interventions which can be adapted to platforms' nuances and implemented with the time remaining before November. Notably, we do not suggest that platforms ban large categories of content or avoid being sites of election information. Nor do we expect that platforms will be able to identify and act upon every piece of election-threatening content created with or published on their surfaces.

There is no way to fully address the digital threats surrounding the U.S.'s 2024 election cycle. Increased scrutiny by the Select Subcommittee on the Weaponization of the Federal Government and via litigation like *Murthy v. Missouri* has put pressure on platforms' integrity measures and coordination with external stakeholders, including government agencies.[3] Moreover, even if platforms were to implement every guardrail and mitigation available, bad actors would still find ways to produce and distribute election-threatening content online.

Recognizing this, **our recommendations offer a pragmatic and systemic approach to risk mitigation — the platform equivalent to an ounce of prevention being worth a pound of cure. They highlight proactive measures that do not censor, but rather lay safeguards along the path to scaled production or distribution.**

We recognize that *neither* silencing voices nor allowing a small set of users to dominate the online environment is healthy for democracy or online communities. And we believe the adoption of these recommendations would meaningfully reduce the volatility the online information environment threatens to inject into the 2024 election. By implementing these measures, platforms will serve as sites for democratic discourse and demonstrate that protecting our experiment in self-government is a priority worth optimizing for.

What's Changed Since 2020

Today's landscape of online platforms is far larger and more fragmented — though no less interconnected — than the field of platforms available during the 2020 election or any previous election cycle. Platforms vary widely in design, content formats, user base, and company size, but broadly fall into three major categories: (1) social media platforms, (2) messaging platforms, and (3) generative AI platforms. None of these platform categories are siloed — instead, their products are complementary and connected, together driving the dynamics of how content is created and spread online.

As the most significant **content *distribution* platforms** in the online ecosystem, **social media and messaging platforms** have been both channels for valuable election information as well as vectors for disinformation and election subversion narratives. **Generative AI platforms**, in contrast, are **content *production* platforms**. Their widespread availability has made it easier than ever to develop high-quality synthetic media across content formats (visual, audio, and text). This includes first and third-party offerings that integrate generative AI capabilities into a range of products and surfaces, including social media and messaging platforms.[4] Examples of synthetic content being created and spread with the aim of influencing elections are already multiplying.[5]

Alongside changes in the platform landscape, the risks facing American elections and democracy more broadly have shifted and escalated since 2020.[6] As a result, the 2024 election faces a rise in threats and harassment directed at election officials,[7] experts' concerns about the increased risk for violence,[8] and narratives proliferating online and offline that erode confidence in our electoral systems.[9]

Social media, messaging and generative AI **platforms inevitably will be critical sources and conduits of election information this cycle. As such, they can and should make choices that will meaningfully enhance the degree to which the 2024 presidential election is free and fair**. Of course, these choices require tradeoffs — both in terms of resource investment and short-term engagement on a platform. Fortunately, there is good precedent for platforms making tough choices that uphold their responsibility as hosts of election information and democratic engagement.

With finite time remaining and platforms' 2024 product roadmaps already in flight, it is now more essential than ever that platforms make these choices again. The work to implement election-protection strategies that are **pragmatic, achievable, and impactful** has begun and continues through Inauguration Day 2025. What follows are recommendations to help achieve those aims.

The Shortlist: Recommendations by Platform Category

Social Media Platforms

Adequate Election Teams Resourcing

Adequately resource election teams, including related Trust and Safety, policy, legal, and operations teams, at least six months before the U.S. general election and maintain this resourcing through Inauguration Day.

Authoritative Voting and Election Information

Amplify accurate, authoritative content on the time, place, and manner of voting and election results for the remainder of the U.S. election season.

Reasonable Usage-Rate Limits

Establish usage-rate limits for inviting, messaging, sharing, commenting, and forwarding features — particularly their usage by accounts and entities that are new, demonstrate suspicious activity, or relate to voting or elections — at least four months prior to the general election through Inauguration Day.

Limiting Distribution of New and Suspicious Accounts and Entities

Limit distribution of content from new accounts and entities as well as accounts and entities that have demonstrated suspicious on-platform activity, at least four months prior to the general election through Inauguration Day.

Messaging Platforms

Adequate Election Teams Resourcing

Adequately resource election teams, including related Trust and Safety, policy, legal, and operations teams, at least six months before the U.S. general election and maintain this resourcing through Inauguration Day.

Authoritative Voting and Election Information

Prominently offer in-product channels, like chatbots, for users to receive authoritative, accurate content on the time, place, and manner of voting and election results for the remainder of U.S. election season.

Reasonable Usage-Rate Limits

Establish usage-rate limits for inviting, messaging, sharing and forwarding features — particularly their usage by accounts and entities that are new or demonstrate suspicious activity — at least four months prior to the general election through Inauguration Day.

Heightened Enforcement on Inauthentic Networks

Prohibit coordinated inauthentic behavior using fake accounts and temporarily reduce the threshold for enforcing on borderline inauthentic account networks, at least four months prior to the general election through Inauguration Day.

Generative AI Platforms

Adequate Election Teams Resourcing

Adequately resource election teams, including related Trust and Safety, policy, legal, and operations teams, at least six months before the U.S. general election and maintain this resourcing through Inauguration Day.

Authoritative Voting and Election Information

Direct users to official sources of accurate, authoritative information on the time, place, and manner of voting and election results for the remainder of U.S. election season.

Disclosing Content Authenticity

Deploy one direct (user-facing) and one indirect (not user-facing) disclosure synthetic-media transparency method for audio and visual synthetic content and conduct public education so diverse audiences and end-users can distinguish AI-generated or modified content.

Election Integrity Policies

Prohibit in API and business policies the use of services or models to interfere with the lawful conduct of elections, including spreading falsehoods concerning election laws or processes or intimidating voters or election officials.

The Shortlist: Cross-Category Platform Recommendations

Read in-depth descriptions on pages 6–11.

	Social Media Platforms	Messaging Platforms	Generative AI Platforms
Adequate Election Teams Resourcing Adequately resource election teams, including related Trust and Safety, policy, legal, and operations teams, at least six months before the U.S. general election and maintain this resourcing through Inauguration Day.	✓	✓	✓
Authoritative Voting and Election Information Prominently offer in-product channels for authoritative, accurate content on the time, place, and manner of voting and election results for the remainder of U.S. election season.	✓	✓	✓
Reasonable Usage-Rate Limits Establish usage-rate limits for inviting, commenting, messaging, sharing and forwarding features — particularly their usage by accounts and entities that are new, demonstrate suspicious on-platform activity, or relate to voting and elections — at least four months before the general election through Inauguration Day.	✓	✓	
Limiting Distribution of New and Suspicious Entities Limit distribution of content from new accounts and entities as well as accounts and entities that have demonstrated suspicious on-platform activity, at least four months prior to the general election through Inauguration Day.	✓		
Heightened Enforcement on Inauthentic Networks Prohibit coordinated inauthentic behavior using fake accounts and temporarily reduce the threshold for enforcing on borderline inauthentic account networks, at least four months prior to the general election through Inauguration Day.		✓	
Disclosing Content Authenticity Deploy one direct (user-facing) and one indirect (not user-facing) disclosure synthetic-media transparency method for audio and visual synthetic content and conduct public education so diverse audiences and end-users can distinguish AI-generated or modified content.			✓
Election Integrity Policies Prohibit in API and business policies the use of services or models to interfere with the lawful conduct of elections, including spreading falsehoods concerning election laws or processes or intimidating voters or election officials.			✓

Recommendations

Platforms can take these steps to safeguard the U.S. election in 2024.

Adequate Election Teams Resourcing Across Platforms

No matter what form election protection takes at a platform, it relies on teams operating with election safety as a top priority leading up to and throughout election season, including Inauguration Day. This can include both internal teams and external partners, like third-party fact checkers or civil society organizations who offer public education on voting and election administration. Internally, these teams vary in size and function across platforms but are typically cross-functional and include fully staffed product teams (product managers, engineers, design and research managers) as well as content policy, partnerships, operations, legal, and communications managers.[10] Across these functions, team members may not all have specific election expertise or be exclusively dedicated to elections. However, at minimum, election leads, particularly in policy and partnerships functions, should be versed on U.S. election administration as well as the specific outlook and risks facing the 2024 cycle.

While individual platforms are best suited to determine what constitutes adequate resourcing for their teams, they should base this assessment on audits that evaluate how a platform could be used to produce or distribute election information.[11] Platforms should prioritize resourcing based on the level of risks across the use cases for creating or spreading election information they identify, especially risks of voter suppression or physical violence. Sufficient resourcing includes staffing, budget, and tooling, including ensuring platforms are able to execute robust on-platform monitoring. Finally, adequate resourcing should account for peak moments in the cycle that will pose elevated risks and require surge capacity and oversight.

Election teams vary in the degree to which they are centralized or dispersed within an organization. Regardless of their form or where they're housed, teams must have a documented understanding of roles — namely, the key decision makers at critical junctures — including amongst a platform's executives and C-suite, legal counsel and operations managers. This understanding should be paired with replicable, documented processes that election teams and decision makers can use to quickly assess and respond to emerging threats. While election teams should engage in thorough red teaming or threat scenario planning to inform their preparations, they will inevitably encounter novel situations during the 2024 cycle. When presented with these situations,

election teams must make difficult decisions in a compressed timeline, which will rely on clear escalation channels and consistent, documented communication.[12]

Authoritative Voting and Election Information Across Platforms

All three categories of platforms should prioritize ensuring their users have consistent access to authoritative information on voting and the 2024 election's administration for the full duration of the cycle, through Inauguration Day. This information should concentrate on information about all stages of voting and election results. Platforms would be wise to rely on partnerships with official election authorities or civil society organizations to equip users with vetted information from authoritative sources.

There are a range of formats and channels that platforms can use to equip users with authoritative election information. For example, social media platforms can amplify such information, whether in-feed or through recommendation surfaces, or prominently display an in-product election hub.[13] Messaging platforms, regardless of whether they employ end-to-end encryption, can ensure users have the option to engage with dedicated chatbots to fact check information or access authoritative election FAQs.[14] Finally, generative AI platforms can direct users to authoritative sources of information in response to relevant queries and at minimum, should train models to refuse to answer election-related queries for which they cannot consistently and accurately provide authoritative information.[15]

Across these delivery mechanisms, platforms should prioritize ensuring that information is accessible to a diverse American audience, including non-English-speaking communities. Platforms should also ensure the information they offer is digestible, timely, and provides sufficient context to help users situate the current moment within the broader electoral process.

Reasonable Usage-Rate Limits for Social Media And Messaging Platforms

Usage-rate limits[16] place a ceiling on the number of times in a certain period any user can employ a specific platform feature like commenting, inviting, messaging, sharing or forwarding. In placing this ceiling, rate limits reduce the likelihood that bad actors, whether relying on bots or prolific human activity, can supercharge distribution of content or entities by abusively overusing a feature. In past U.S. elections, there has been a recurring dynamic of a small set of superusers having a significantly outsized role in producing and spreading harmful election-related content, including disinformation and calls for political violence.[17] These superusers have illustrated that social media and messaging platforms offer features that, when used at extreme outlier or "spammy" levels, can be vectors for manipulation.

This recommendation suggests platforms implement rate limits that *narrowly* prevent extreme overuse. Establishing a reasonable, focused threshold for rate limits requires platforms to carefully balance tradeoffs with on-platform engagement while also recognizing how rate limits will impact both legitimate and manipulative usage of a feature. Actual implementation will vary by platform, but in practice, successful deployment would mean that a platform sets a targeted rate limit that only affects a small sliver of users' "spammy" activity.[18] What's more, rate limits do not ban accounts from ever using a feature — they prevent outlier usage for a defined duration, after which point an account can begin using that feature again.[19]

Platforms should consider how rate limits should be adjusted in response to the increased risks and dynamic nature of the election environment. For example, a platform could apply a rate limit in a targeted manner prior to voting and broaden its application once voting begins, continuing to adjust it as needed for higher-risk periods or significant fluctuations in platform usage. In addition, platforms should diligently apply rate limits to categories of content or entities that likely pose higher risks during election season, such as new accounts and entities,[20] accounts and entities that have demonstrated suspicious on-platform behavior,[21] or accounts and entities that relate to voting or elections. At minimum, platforms should plan for aggressive applications of rate limits as a break-the-glass measure and have clear documentation for the criteria that would trigger this deployment.

The rate limit recommendations offered for social media and messaging platforms differ in two respects. First, social media platforms typically offer a commenting feature absent on messaging platforms. Second, a number of messaging platforms in the U.S. offer end-to-end encryption. Where social media platforms can employ algorithmic classifiers to distinguish and categorize content, encrypted messaging platforms do not view the content shared on their platforms, and thus cannot distinguish among categories of content distributed on their surfaces. As a result, the social media platform recommendation suggests applying rate limits to election and voting-related content and entities, as defined by individual platforms. By comparison, our recommendation for messaging platforms recognizes that rate limits can't be applied based on a category of content on encrypted channels.[22]

Limiting Distribution of New and Suspicious Accounts And Entities for Social Media Platforms

Distribution on a social media platform relies on algorithmically ranking content. Each platform employs its own set of ranking systems and criteria, but largely they function in a similar manner to transform what would be an impossibly overwhelming volume of content into a functional, curated feed or list for users.[23] Platforms broadly optimize their ranking systems to deliver to each user a unique set of content based on what

delivers the highest value to the company, which most platforms define as on-platform engagement.[24]

Legacy social media platforms each also employ and monitor on-platform signals to identify suspicious or unusual activity. This can include outlier levels of activity or growth, particularly after periods of account inactivity, as well as tracking specific policy violations associated with an account or entity. Commonly, monitoring also looks for spam-like activity, which platforms widely recognize as behavior that should be curtailed. In executing on-platform monitoring for any of these signals, platforms should especially prioritize accounts that have desirable characteristics, such as having verified status.

In addition to demonstrating how hyperactive users have proven to be recurring spreaders of election-threatening narratives, past election cycles have highlighted how new accounts and entities, particularly those that gain viral traction and growth, can be used to publish and spread election-threatening content.[25] As a result, Trust and Safety teams at legacy social media platforms have included heightened safeguards on newly-created entities or accounts as break-the-glass measures, including limiting invitations to join or follow new entities or avoiding recommending content from new entities or accounts.[26]

During the sensitive period of the 2024 election cycle, platforms should limit the distribution of content from both new accounts and entities as well as those that have signaled suspicious on-platform activity. Platforms are best suited to determine what constitutes a new or suspicious account or entity. In doing so, they should consider not only on-platform behaviors that signal suspicion, but those that suggest an account is legitimate or trustworthy. Accounting for such signals in determining a new account's trustworthiness can help ensure that new, legitimate accounts are not indefinitely placed at a distribution disadvantage.

Heightened Enforcement on Inauthentic Networks for Messaging Platforms

Coordinated networks of fake accounts or bots have been a hallmark of influence operations during past election cycles, including those led by foreign actors.[27] Recent self-published reports from platforms demonstrate the extent to which this tactic is still in use by foreign actors on distribution platforms.[28]

In addition, researchers and monitors have highlighted how the widespread availability of generative AI has made managing those networks easier and more convincingly human than ever before.[29] In recognition of the new state of play, it is essential that messaging platforms, regardless of whether they are end-to-end encrypted, have policies that prohibit inauthentic behavior, and specifically the coordinated use of fake accounts or entities.[30] Platforms, both social media and messaging, who have adopted policies like these self report that resulting investigations that focus on account

behavior, rather than content, have created resiliency to threat actors attempting to use synthetic content in covert influence operations [31]

In addition, starting at least four months prior to the U.S. general election through Inauguration Day, messaging platforms should reduce the threshold at which they take action on suspected inauthentic account networks. These thresholds should employ behavioral signals that can be identified even on encrypted platforms, such as unusual spikes in account or messaging activity or rates of activity inconsistent with a human user (i.e., the rate at which messages are sent or typed).[32] Platforms may also consider, where resources permit, training AI models to detect coordinated inauthentic behavior, using on-platform data to compare past and recent behavior of inauthentic account networks with the activity of typical human users.[33] Recognizing that broadened enforcement may result in false positives, platforms should offer users in-product appeals channels to request platforms review enforcement decisions, as appropriate.

Platforms are best positioned to evaluate and set thresholds in a way that accounts for heightened risks around the 2024 cycle and the new capabilities of AI-enabled networks. They should monitor and adjust these thresholds throughout election season to respond to evolving online dynamics. Finally, as the implications of generative AI's usage by threat actors is evolving — including how foreign actors will use the technology — platforms should exchange information between each other to identify cross-platform influence operations.[34]

Disclosing Content Authenticity for Generative AI Platforms

The anticipated proliferation of synthetic content in the U.S.'s election information ecosystem will require audiences, journalists, and distribution platforms, like social media and messaging platforms, to grapple in new ways with content authenticity. While not a silver bullet, generative AI platforms should employ synthetic media transparency methods,[35] both direct (user facing) and indirect (not user facing) disclosure methods, for their visual and audio content.[36] Alongside these disclosure methods, generative AI platforms should adopt policies that prohibit users from representing the output of a generative AI platform as *not* synthetic, which should apply to first and third-party usage of models.[37]

There is not one form of synthetic media transparency that alone can address the challenges introduced by generative AI's widespread availability. Therefore, we believe platforms should take a balanced, portfolio approach to disclosure. At minimum, platforms should employ at least one synthetic media transparency method that provides direct disclosure to end users to signal content that is AI-generated or AI-modified. This disclosure can take the form of content labels or overlays such as visible watermarks,[38] but should be designed for the general public's comprehension.

Unfortunately, direct disclosure methods, like visible watermarking, are unlikely to withstand bad actors' circumvention. As a result, generative AI platforms should also implement at least one indirect disclosure method for their audio and visual content, such as signed metadata or invisible watermarks. Rather than being user-facing, indirect disclosure methods signal to entities involved in contents' development and distribution — such as social media and messaging platforms — when a piece of content is AI-generated or modified.

As generative AI platforms adopt direct disclosure synthetic media transparency methods, they should help audiences and end-users understand those disclosures' significance.[39] No matter their form, these methods are new to the American public and robust digital literacy campaigns should accompany them. Platforms can use a combination of approaches including funding programs with trusted intermediaries, in-product education, and cross-industry partnerships to educate voters.

Election Integrity Policies for Generative AI Platforms

Legacy social media and messaging platforms have experienced one or more U.S. election cycles, but 2024 will be a testing ground for more recently launched generative AI platforms. As yet, generative AI platforms largely lack election-specific terms of service or usage policies analogous to those social media and messaging platforms have on the books.[40]

Having election-specific policies in place ensures generative AI platforms clearly and publicly convey the behaviors they will monitor and enforce. Naming election-related prohibited applications clarifies whether, for example, voter suppression or election subversion efforts will qualify under broad policies prohibiting "harmful" or "misleading" content.[41] The recently announced Tech Accord to Combat Deceptive Use of AI in 2024 Elections has acknowledged the importance of "providing transparency to the public...by publishing the policies that explain how we will address such content."[42] This is critical for platforms' API or business service terms because abusing these offerings can result in the production and distribution at scale of election-threatening synthetic content.[43] The election-specific policies we propose (bans on falsehoods concerning election laws, processes, or procedures and intimidating voters or election officials) are also consistent with U.S. law, which includes numerous provisions prohibiting interference with the right to vote and voter intimidation.[44]

Notes

1 Editorial, *This Isn't Just an Election Year. It's the Year of Elections*, WASH. POST (Dec. 31, 2023, 7:15 AM), https://www.washingtonpost.com/opinions/2023/12/31/elections-world-presidents-legitimacy.

2 *See Elections Everywhere All at Once*, ATLANTIC COUNCIL (Dec. 6, 2023, 3:00 PM), https://www.atlanticcouncil.org/event/elections-everywhere-all-at-once; Katie Harbath & Ana Khizanishvili, *Insights from Data: What the Numbers Tell Us About Elections and Future of Democracy*, INTEGRITY INSTITUTE (Oct. 30, 2023), https://integrityinstitute.org/blog/insights-from-data.

3 *See* Emily Brooks & Rebecca Klar, '*Weaponization' Subcommittee Members Spar Over 'Twitter Files,'* THE HILL (March 9, 2023, 2:21 PM), https://thehill.com/homenews/3892219-weaponization-subcommittee-members-spar-over-twitter-files; Ann E. Marimow, *Supreme Court Says White House May Continue Requests to Tech Companies*, WASH. POST (Oct. 20, 2023, 6:14 PM), https://www.washingtonpost.com/politics/2023/10/20/supreme-court-tech-companies-social-media-posts/.

4 For example, Meta has announced their integration of generative AI across Facebook, Instagram, Messenger, and WhatsApp, including Meta AI. *What's New Across Our AI Experiences*, META: NEWSROOM (Dec. 12, 2023, 12:05 PM), https://about.fb.com/news/2023/12/meta-ai-updates.

5 *See* Morgan Meaker, *Slovakia's Election Deepfakes Show AI is a Danger to Democracy*, WIRED (Oct. 3, 2023, 7:00 AM), https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy; Ali Swenson & Will Weissert, *New Hampshire Investigating Fake Biden Robocall Meant to Discourage Voters Ahead of Primary*, AP NEWS (Jan. 22, 2024, 11:32 PM), https://apnews.com/article/new-hampshire-primary-biden-ai-deepfake-robocall-f3469ceb6dd613079092279946 63db5; Rishi Iyengar, *How China Exploited Taiwan's Election — and What It Could Do Next*, FOREIGN POLICY (Jan. 23, 2024, 2:37 PM), https://foreignpolicy.com/2024/01/23/taiwan-election-china-disinformation-influence-interference; Kate Lamb, Fanny Potkin & Ananda Teresia, *Generative AI May Change Elections This Year. Indonesia Shows How*, REUTERS (Feb. 8, 2024, 6:59 AM), https://www.reuters.com/technology/generative-ai-faces-major-test-indonesia-holds-largest-election-since -boom-2024-02-08.

6 Deana El-Mallawany, Justin Florence & Jessica Marsden, *The Triple Threat to Democracy in 2024*, PROTECT DEMOCRACY (Jan. 10, 2024), https://protectdemocracy.org/work/the-triple-threat-to-democracy-in-2024.

7 William Brangham & Ian Couzens, *Election Workers Face Violent Threats and Harassment Amid Dangerous Political Rhetoric*, PBS NEWSHOUR (Nov. 16, 2023, 6:35 PM), https://www.pbs.org/newshour/show/election-workers-face-violent-threats-and-harassment-amid-dangerous-political-rhetoric.

8 Jennifer Dresden & Lilliana Mason, *Violence and Democracy Impact Tracker: December 2023 Update*, PROTECT DEMOCRACY (Dec. 12, 2023), https://protectdemocracy.org/wp-content/uploads/2023/12/23.12.12-VDIT-Q4Update-1.pdf.

9 El-Mallawany, *supra* note 6.

10 *Elections Integrity Best Practices, Elections Integrity Series Part 1*, INTEGRITY INSTITUTE (May 17, 2023), https://integrityinstitute.org/s/Final-Elections-Best-Practices-Guide-Part-1_2023-05-24.pdf.

11 *Id.*

12 *Id.*

13 In 2020, Twitter introduced an election hub "to help Americans prepare for the most uncertain election in modern U.S. history." Taylor Hatmaker, *Twitter Debuts U.S. Election Hub to Help People Navigate Voting in 2020*, TECHCRUNCH (Sept. 15, 2020, 1:00 PM), https://techcrunch.com/2020/09/15/twitter-election-hub-voting-tools.

14 One example of a chatbot-enabled, fact-checking tipline program on an encrypted messaging platform is Meedan's election fact-checking programs on WhatsApp. *Elections: Verified Content for the Voting Public*, MEEDAN (last visited Feb. 27, 2023), https://meedan.com/programs/elections.

15 OpenAI announced a partnership with the National Association of Secretaries of State that will ensure ChatGPT users are directed to CanIVote.org if they pose election-related procedure questions. OpenAI, *How OpenAI is Approaching 2024 Worldwide Elections*, OPENAI (Jan. 15, 2024), https://openai.com/blog/how-openai-is-approaching-2024-worldwide-elections#OpenAI.

16 *See* Ravi Iyer, *A Concise Social Media Design Election Advocacy Guide for 2024*, DESIGNING T*OMORROW* (Jan. 19, 2024), https://open.substack.com/pub/psychoftech/p/a-concise-social-media-design-election?r=2b7wo9&utm_campaign=post&utm_medium=email.

17 For example, an internal analysis at Facebook determined that 0.3% of users were responsible for 30% of the group invites that resulted in the original "Stop the Steal" Facebook group growing to 360,000 members in 24 hours, with 2.1 million membership requests still pending when it was taken down. Similarly, internal research at Facebook found that one individual issued 400,000 invitations to QAnon groups in six months. *See* JEFF HORWITZ, BROKEN CODE 205, 219 (2023).

18 As described in *Broken Code,* the internal team at Facebook created to fight Dedicated Vaccine Discouragement Entities "set the goal of limiting the anti-vax activity of the top .001 percent of users — a group that turned out to have a meaningful effect on overall discourse." *Id.* at 247.

19 For example, if a rate limit establishes a threshold such that no user can send no more than 50 invitations to a group each day, an impacted user would not be able to send their 51st invitation in that twenty-four hour period. Once the defined time period — a speed bump, so to speak — for the rate limit has passed, the user would be able to again send invitations to the group.

20 *See infra* pp. 12-13.

21 *See infra* p. 12.

22 WhatsApp, an encrypted messaging platform, employs rate limits to maintain the private nature of their service *and* safeguard elections. In 2019, WhatsApp set a content-level rate limit by restricting message forwarding to five chats at a time. In addition, WhatsApp separately limited the reforwarding of viral messages. Specifically, the platform labeled messages that had been reforwarded many times and limited their resharing to one chat at a time. Finally, in recognition of bad actors with political motivations, WhatsApp also maintained account-level rate limits on the number of groups an account could create within a specific time period. *More Changes to Forwarding,* WHATSAPP (Jan. 21, 2019), https://blog.whatsapp.com/more-changes-to-forwarding; *About WhatsApp and Elections,* WHATSAPP, https://faq.whatsapp.com/518562649771533/ ?helpref=uf_share (last visited Feb. 28, 2024); *Stopping Abuse: How WhatsApp Fights Bulk Messaging and Automated Behavior,* WHATSAPP (Feb. 6, 2019), https://scontent-sjc3-1.xx.fbcdn.net/v/t39.8562-6/299911313_583606040085749_3003238759000179053_n.pdf?_nc_cat=101&ccb=1-7&_nc_sid=b8d81d&_nc_ohc=_NqFMccy7U4AX8SrXhF&_nc_ht=scontent-sjc3-1.xx&oh=00_AfAUg7YY5qSKrwBdzx9Y-pl0_e87YD89fXfBPiRwjrycmQ&oe=65E50694.

23 Brief of the Integrity Institute and Algotransparency as Amici Curiae in Support of Neither Party, Gonzalez, et. al, v. Google LLC., 598 U.S. ____ (2023), https://www.supremecourt.gov/DocketPDF/21/21-1333/249279/20221207100038897_21-1333_Amici%20Brief.pdf.

24 *Id.*

25 HORWITZ, *supra* note 17.

26 *Id.* at 213.

27 S. Rep. No. 166-290 at 18 (2020).

28 *See* Ben Nimmo, et. al., *Third Quarter Adversarial Threat Report,* META (November 2023), https://scontent-sjc3-1.xx.fbcdn.net/v/t39.8562-6/406961197_3573768156197610_1503341237955279091_n.pdf?_nc_cat=105&ccb=1-7&_nc_sid=b8d81d&_nc_ohc=ov1yoGD30OsAX9iqZSy&_nc_ht=scontent-sjc3-1.xx&oh=00_AfCIj9mm7ATOzgdpUm22xhiMZ8GfUvIkYgS5jkVal1ae-Q&oe=65E372D2.

29 Though AI-generated fake profile pictures have been used since 2019, inauthentic accounts then relied largely on human labor, often in the form of troll farms. This meant that inauthentic networks could be detected by identifying patterns in both account behaviors and content, such as suspiciously coordinated messaging schedules, or frequently repeated phrases or spelling errors. In the era of generative AI, inauthentic networks can now be managed at scale, while avoiding some of these common signals of suspicious activity. For example, generative AI can be used to create many variations of the same message, while largely avoiding repetitive phrasing and spelling errors. In addition, AI chatbots have significantly changed the degree to which bad actors need human labor to manage an influence operation. *See* William Marcellino, et. al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*, RAND CORPORATION (Sept. 7, 2023), https://www.rand.org/pubs/perspectives/PEA2679-1.html.

30 Facebook's Inauthentic Behavior Policy (which applies to messaging platforms Messenger and Instagram) is an example of such a policy. The policy specifically prohibits "Coordinated Inauthentic Behavior," which is defined as "the use of multiple Facebook or Instagram assets, working in concert to engage in Inauthentic Behavior … where the use of fake accounts is central to the operation." *Inauthentic Behavior*, META (April 25, 2022), https://transparency.fb.com/policies/community-standards/inauthentic-behavior.

31 Nimmo, *supra* note 28 at 26.

32 Stopping Abuse*, supra* note 22 at 7.

33 *Id.*

34 *See* Nimmo, *supra* note 28 at 17.

35 A growing number of terms are used to describe strategies to disclose whether content is synthetic. Here, synthetic media transparency methods is defined using the Partnership on AI's Glossary for Synthetic Media Transparency Methods as "[t]he umbrella term used to describe signals for conveying whether a piece of media is AI-generated or AI-modified." PAI Staff, *Building a Glossary for Synthetic Media Transparency Methods, Part 1: Indirect Disclosure*, PARTNERSHIP ON AI (Dec. 19, 2023), https://partnershiponai.org/glossary-for-synthetic-media-transparency-methods-part-1-indirect-disclosure.

36 The recently announced Tech Accord to Combat Deceptive Use of AI in 2024 Elections includes provenance as one of its seven principal goals. *A Tech Accord to Combat Deceptive Use of AI in 2024 Elections*, AI ELECTIONS ACCORD (Feb. 16, 2024), https://www.aielectionsaccord.com/uploads/2024/02/A-Tech-Accord-to-Combat-Deceptive-Use-of-AI-in-2024-Elections.FINAL_.pdf.

37 For example, Google's Generative AI Prohibited Use Policy prohibits the "misrepresentation of the provenance of generated content" created by relevant Google services "by claiming content was created by a human … in order to deceive." *Generative AI Prohibited Use Policy*, GOOGLE (March 14, 2023), https://policies.google.com/terms/generative-ai/use-policy.

38 PAI Staff *supra* note 35.

39 The recently announced Tech Accord to Combat Deceptive Use of AI in 2024 Elections includes Public Awareness as one of its seven principal goals, specifically, "Engaging in shared efforts to educate the public about media literacy best practices, in particular regarding Deceptive AI Election Content, and ways citizens can protect themselves from being manipulated or deceived by this content." AI ELECTIONS ACCORD, *supra* note 36.

40 *See, e.g., Misinformation Policy Explainer*, DISCORD (Oct. 24, 2023), https://discord.com/safety/misinformation-policy-explainer (prohibiting misinformation about "the integrity of a civic process — specifically, around issues that could delegitimize results or undermine faith in public institutions"); *Civic and Election Integrity,* TIKTOK (March 2023), https://www.tiktok.com/community-guidelines/ en/integrity-authenticity/#2 (prohibiting the misinformation about the "laws, processes, and procedures that govern the organization and implementation of elections … " as well as misinformation about the outcome of an election).

41 Speechify's Prohibited Uses of the Service, for example, includes the use of "the Services for any illegal, immoral or harmful purpose." *Terms & Conditions*, SPEECHIFY (May 25, 2023), https://speechify.com/terms.

42 AI ELECTIONS ACCORD, *supra* note 36.

43 Midjourney's Terms of Service offers an example of a generative AI platform that has election-specific policy language in place, specifically prohibiting users from using the service "to try to influence the outcome of an election." *Terms of Service*, MIDJOURNEY (Dec. 22, 2023), https://docs.midjourney.com/docs/terms-of-service.

44 *See, e.g.*, 18 U.S.C. § 241 (criminalizing interference with the right to vote); *United States v. Mackey*, 652 F. Supp. 3d 309 (E.D.N.Y. 2023) (§ 241 applies to a scheme to distribute false information about voting by text); 42 U.S.C. § 1985 (imposing civil liability for conspiracies to intimidate voters in federal elections).

Protect Democracy is a nonpartisan nonprofit organization dedicated to preventing American democracy from declining into a more authoritarian form of government.

protectdemocracy.org