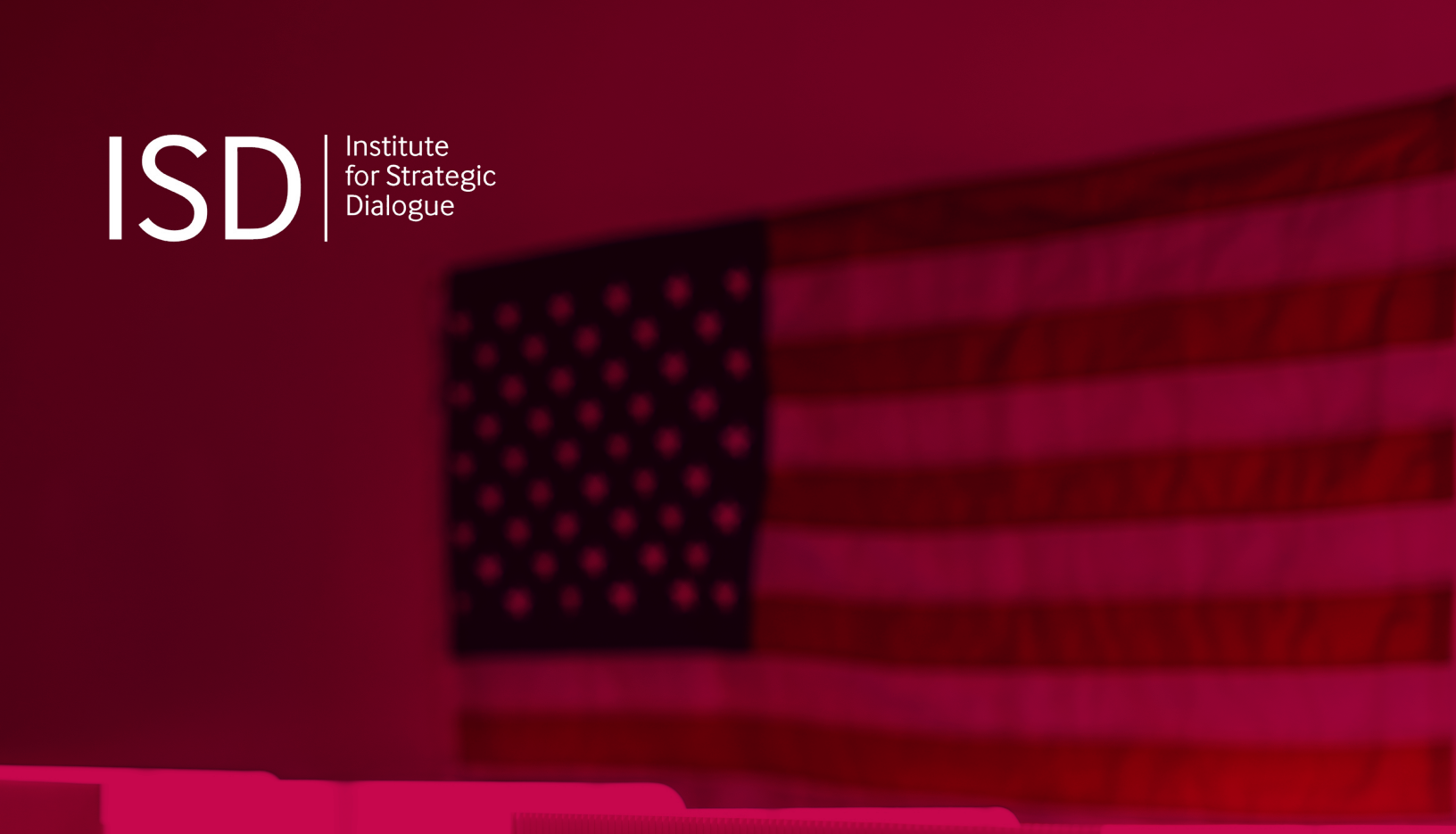
ISD | Institute for Strategic Dialogue





US Election Platform Preparedness 2026

Hannah Wallinger, Valeria de la Fuente, Samantha Fagone and Max Read

ISD | Institute for Strategic Dialogue

Amman | Berlin | London | Paris | Toronto | Washington DC

Copyright © Institute for Strategic Dialogue (2026). ISD-US is a non-profit corporation with 501(c)(3) status registered in the District of Columbia with tax identification number 27-1282489. Details of the Board of Directors can be found at www.isdglobal.org/isd-board. All Rights Reserved.

www.isdglobal.org

Contents

Overview 4

US Election Platform Preparedness: Comparative Analysis 5

Key Risks Identified 12

In-Depth Analysis 14

- AI chatbots are now standard across every platform, with limited transparency around election-specific safeguards 14
- Every major platform has expanded its in-house AI generation tools since 2024, but protections for political figures remain inconsistent 14
- Platforms lack dedicated policies for AI-generated political content 15
- Only Meta and YouTube have built AI-specific disclosure rules into their political ad products 15
- AI labeling may depend on platforms' own tools or user self-disclosure 16
- Platforms are shifting from third-party fact-checking toward crowd-sourced moderation 16
- AI-powered moderation is expanding faster than transparency around its use 17

Conclusion 18

Overview

The 2026 midterms are the first US election in which AI is built into every major social media platform, from embedded chatbots and media generation tools to content moderation.

Ahead of November, ISD assessed how platforms including Meta-owned Facebook and Instagram, X, YouTube, Snapchat and TikTok are preparing to protect free and open political discourse, election integrity and public safety through their policies, public commitments and product features. The assessment also examines how platform preparedness has changed since our 2024 assessment and comparative analysis.

This report's central finding is that transparency has not kept pace with AI integration. Many AI features and integrations are new, with some launching only months before the election, yet platforms have disclosed little about how they work, how they are tested or the extent to which they have replaced human oversight. Platforms provide researchers with little to no access to these systems, and any access granted is voluntary. In the EU, by contrast, the Digital Services Act gives vetted researchers a legal right to data. Without a US framework requiring platforms to disclose information or conduct risk assessments, there is little incentive for platforms to open these systems to public scrutiny.

At the same time, platforms are reducing their reliance on professional human oversight. Third-party fact-checking, as well as trust and safety teams, have been scaled back in favor of crowd-sourced and automated systems. These approaches can offer important benefits, including operating at greater scale, but relying on them as the primary means of oversight during a fast-moving election cycle remains an untested proposition.

Taken together, these trends point to a gap in platform preparedness. AI is becoming a central voter perception and behavior, while transparency and independent scrutiny have not kept pace. As a result, voters, election officials and researchers have limited visibility into how platforms will respond when election-related risks emerge.

US Election Platform Preparedness: Comparative Analysis

Note: Bolded cells indicate a rating that has changed since 2024.

Information Integrity (User-Generated Content)

	Meta		X		YouTube		Snapchat		TikTok	
	2026	2024	2026	2024	2026	2024	2026	2024	2026	2024
1. Does the platform remove (or downrank) content that includes false or misleading information on how to participate in an election (includes times, dates, methods and eligibility requirements)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Does that platform remove (or downrank) content that falsely asserts there will be widespread election fraud or vote counting errors?	**Partial** [1]	Yes	Partial [2]	Partial	Yes	Yes	Yes	Yes	Yes	Yes
3. Does the platform remove (or downrank) content that contains false or misleading information on candidate eligibility requirements?	Partial	Partial	Partial	Partial	Yes	Yes	Partial	Partial	Yes	Yes
4. Does the platform remove false claims regarding the 2020 election results or redirect users to accurate information about the 2020 election results?	Partial	Partial	No	No	Partial	Partial	Yes	Yes	Yes	Yes
5. Will the platforms remove or downrank content which claims victory before the election is called?	Partial	Partial	Unclear	Unclear	Unclear	Unclear	Yes	Yes	Yes	Yes

Notes

1 Meta ended its US third-party fact-checking program in 2025 and now uses a crowdsourced Community Notes system.

2 X's Civic Integrity policy relies on Community Notes, which do not trigger removal, labels or reduced distribution, alongside a narrow Civic Integrity Label for critical escalations.

6. Is the platform partnering with external experts to identify coordinated influence operations? Have details on the scope and results of those partnerships been publicly disclosed?	No / Unclear	No / Unclear	No / Unclear	No / Unclear	No / Unclear	No / Unclear	No / Unclear	No / Unclear	**Yes / No**[3]	No / Unclear
7. Are there policy exemptions (or instances where EDSA exceptions may be made) for political figures and candidates?	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes
8. Is the platform partnered with third party fact-checkers? If so, what are the turnaround times?	**No (US)**	Yes / Unclear	No / N/A	No / N/A	**Partial / Unclear**	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear
9. Does the violation of misinformation policies related to the election or civic integrity result in account suspension? If so, after how many violations?	Yes / Unclear	Yes / Unclear	Unclear / Unclear	Unclear / Unclear	Yes / 3 strikes	Yes / 3 strikes	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear
10. Will the platform be promoting verified, trusted election information from third-party sources to platform users?	Yes	Yes	**No**[4]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
11. Are the platform's integrity and authenticity policies applied to livestreaming product features (both audio and audiovisual)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes

3 TikTok states that it collaborates with external intelligence vendors on specific investigations, and its monthly covert influence operations reports record whether a network was detected internally or through an external report. TikTok does not name partners or disclose the scope of these collaborations.

4 X's October 2026 midterm announcement describes its Civic Integrity rules, a reporting option that routes reports into Community Notes requests, and labels in urgent cases, but does not describe directing users to official voting information. In August 2024, X directed Grok election queries to Vote.gov; ISD found no documentation that this practice remains in place.

12. Does the platform remove or downrank synthetically generated or misleadingly edited content of candidates, elected officials or individuals involved in the administration of elections, that could be reasonably judged to cause harm to those individuals?	Partial	Partial	Partial	Partial	Partial	Partial	Yes	Yes	**Partial**	Unclear

Political Ads

	Meta		X		YouTube		Snapchat		TikTok	
	2026	2024	2026	2024	2026	2024	2026	2024	2026	2024
13. Does the platform allow political ads?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
14. If so, are political ads subject to the same or more stringent content policies regarding misinformation?	More Stringent	More Stringent	Equally Stringent	Equally Stringent	Equally Stringent	Equally Stringent	Equally Stringent	Equally Stringent	N/A	N/A
15. Are political ads eligible for third party fact-checking?	No	No	No	No	Unclear	Unclear	Yes	Yes	N/A	N/A
16. Are political ads required to disclose funding sources?	Yes	Yes	Partial	Partial	Yes	Yes	Yes	Yes	N/A	N/A
17. Are political ads subject to the same or more stringent content policies, specifically regarding hate and extremist content?	More Stringent	More Stringent	Equally Stringent	Equally Stringent	Equally Stringent	Equally Stringent	Equally Stringent	Equally Stringent	N/A	N/A
18. Are political ads permitted to use synthetic media or misleadingly edited media?	Yes (requires disclosure)	Yes (requires disclosure)	Yes	Yes	Yes (requires disclosure)	Yes (requires disclosure)	No	No	N/A	N/A
19. Is there a publicly available ads library?	Yes	Yes	Partial	Partial	Yes	Yes	Partial	Partial	No	No
20. Are content creators required to disclose when political content has been sponsored and who has sponsored the content?	Yes	Yes	Partial	Partial	Partial	Partial	N/A	N/A	N/A	N/A

Hate Speech and Violent Extremism (User-Generated Content)

	Meta		X		YouTube		Snapchat		TikTok	
	2026	2024	2026	2024	2026	2024	2026	2024	2026	2024
21. Is the platform putting in place prevention measures for domestic hate and extremism, such as product interventions to disincentivize hateful behavior?	Partial[5]	Partial	No	No	Partial	Partial	No	No	Partial	Partial
22. Is the company working/partnering with external experts to support in proactively identifying hate speech and violent extremist content on its platform/s and corresponding policy gaps throughout the election cycle?	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear
23. Are there policies against promotion or incitement of election violence, militarization and calls for intimidation (e.g., bringing weapons to polling stations)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
24. Does the platform remove (or downrank) content which includes violent threats?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
25. Are there any measures in place to specifically protect candidates and election officials from targeted hate speech, harassment or threats on the platforms?	Partial	Partial	No	No	**Partial**[6]	No	No	No	**Partial**	No
26. Are the platform's hate speech, violence and extremism policies applied to livestreaming product features (both audio and audiovisual)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes

5 In January 2025, Meta announced it would remove restrictions on topics including immigration and gender identity and focus proactive enforcement on illegal and high-severity violations.

6 In September 2026, Google announced tools for candidates and journalists to request removal of AI imagery featuring their likeness on YouTube.

27. Are the platform's hate speech, violence and extremism policies applied to comment features?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
28. Does the platform remove or downrank content that calls for or incites coordinated harassment (both online and offline)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Resourcing (Internal and External)

	Meta		X		YouTube		Snapchat		TikTok	
	2026	2024	2026	2024	2026	2024	2026	2024	2026	2024
29. Is there a dedicated US-based election integrity team, if so, how large is that team?	No / Unclear	No / Unclear	No / Unclear	No / Unclear	Yes / Unclear	Yes / Unclear	No / N/A	No / N/A	Yes	Yes
30. Are Spanish speakers included on that team?	N/A	N/A	N/A	N/A	Unclear	Unclear	N/A	N/A	Yes	Yes
31. Have there been staffing cuts to election and civic integrity teams since the last federal election (2022/2024)?	Yes	Yes	Yes	Yes	Unclear	Unclear	**Unclear**[7]	No	**Unclear**[8]	No
32. Is there a viral content review system in place? If so, are minimum thresholds or definitions for virality disclosed?	Yes / Unclear	Yes / Unclear	Yes / Unclear	Yes / Unclear	No / Unclear	No / Unclear	No / Unclear	No / Unclear	Yes / Unclear	Yes / Unclear

Transparency

	Meta		X		YouTube		Snapchat		TikTok	
	2026	2024	2026	2024	2026	2024	2026	2024	2026	2024
33. Will the platform be releasing findings on identified disinformation networks on an ongoing basis (monthly or more) throughout the election cycle?	No	No	No	No	No	No	No	No	Yes	Yes

Notes

7 Snapchat cut about 16 percent of its workforce, roughly 1,000 jobs, in April 2026, citing AI efficiencies. Reports did not identify which teams were affected, and ISD found no information on Snapchat's election integrity staffing.

8 Reuters reported layoffs in February 2025, and the Financial Times reported further cuts in August 2025 as TikTok shifted more moderation to AI. None of the reports says whether election integrity teams were affected.

34. Will findings on malicious actors and/or harmful content trends relevant to civic discourse be shared on an ongoing basis (monthly or more) during the election cycle?	No	No	No	No	No	No	No	No	Yes	Yes
35. Have civil society and academic researchers been provided API access for the purposes of independent election integrity research?	Partial	Partial	Partial	Partial	Partial	Partial	No	No	Partial	Partial

State-Affiliated Media

	Meta		X		YouTube		Snapchat		TikTok	
	2026	2024	2026	2024	2026	2024	2026	2024	2026	2024
36. Does the platform clearly label state-affiliated media at the post/video level?	Yes	Yes	**No**[9]	Yes	Yes	Yes	No	No	Yes	Yes
37. Does the platform clearly label state-affiliated media at the account level?	Yes	Yes	**No**	Yes	No	No	No	No	Yes	Yes
38. Does the platform prohibit state-affiliated media from advertising within the US?	Yes	Yes	Yes	Yes	No	No	Partial	Partial	Yes	Yes
39. If yes, does the platform label such ads?	N/A	N/A	N/A	N/A	Yes	Yes	Unclear	Unclear	N/A	N/A

AI Integrations

	Meta		X		YouTube		Snapchat		TikTok	
	2026	2024	2026	2024	2026	2024	2026	2024	2026	2024
40. Does the platform have an integrated chatbot feature?	Yes	Yes	**Yes**[10]	Partial	**Yes**[11]	Partial	Yes	Yes	**Yes**[12]	Partial
41. If yes, is the platform partnering with external organizations to provide reliable election information?	Yes	Yes	**No**	Yes	Not doc.	Not doc.	Not doc.	Not doc.	Not doc.	Not doc.

Notes

9 X, then Twitter, removed its state-affiliated and government-funded media labels in April 2023.

10 Grok became available at no cost to all X users in December 2024.

11 Ask YouTube is open to all signed-in US users.

12 TikTok rolled Tako out widely in February 2026.

42. Does the platform have an integrated media generation or editing feature?	Yes	Yes	**Yes**[13]	Partial	Yes	Yes	Yes	Yes	Yes	Yes
43. If yes, does the platform have dedicated policies governing generation or editing of political content?	Yes	Yes	No	No	**Partial**	No	Yes	Yes	Partial	Partial
44. Are public figures or candidates treated differently under these policies?	Yes	Yes	No	No	**Partial**	No	Yes	Yes	Yes	Yes
45. Does the platform publicly state that LLMs play a substantial role in content moderation?	**Yes**	No	**Partial**	No	Partial	Partial	No	No	**Yes**	No
46. Does the platform publicly document election-specific AI safeguards or limitations?	**Yes**[14]	No	No	No	No	No	Yes	Yes	**Partial**[15]	No
47. Does the platform require labeling of AI-generated content, including AI-generated political content?	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
48. If yes, is labeling automatic, user-provided or both?	Both	Both	N/A	N/A	**Both**[16]	User-provided	N/A	N/A	Both	Both
49. Do content reporting features give users the option to report deceptive AI-generated or manipulated political content? (As in, is there a dedicated AI/deepfake reporting category?)	No	No	No	No	Partial	Partial	No	No	Yes	Yes
50. Does the platform have AI-specific policies or disclosure requirements for political advertising?	Yes	Yes	Partial	Partial	Yes	Yes	Partial	Partial	N/A	N/A

Notes

13 xAI introduced Aurora, Grok's image generator, to all users in December 2024.

14 In September 2026, a Meta spokesperson stated that Meta AI provides users asking voting questions with local information or directs them to government sources.

15 TikTok's October 2026 midterm announcement promises more monitoring of misleading AI-generated content.

16 In May 2026, YouTube stated that it will automatically label undisclosed photorealistic AI content when its systems detect it. In March 2024, YouTube had stated that it only "may add a label" in some cases.

Key Risks Identified

Platforms have offered little clarity on how their AI chatbots will handle election-related queries. AI chatbots are now built into all major social media platforms. For many users, chatbots may be a source of election-related information. Yet no platform has disclosed how its chatbot will handle election questions. ISD testing in June 2026 found that leading AI models, including X's Grok and Meta's Muse Spark, frequently gave inaccurate or incomplete answers to basic voter-information questions, with some responses potentially misleading or discouraging eligible voters from voting. Without greater transparency about how these tools handle election questions, errors may go unnoticed until it is too late to correct them.

Policies governing public officials and AI are fragmented and incomplete. Meta and Snapchat have stated that they restrict their AI image tools from generating political figures, and TikTok applies a similar restriction within its advertiser-facing Symphony Creative Studio. Meanwhile, xAI's Acceptable Use Policy does not specify any such restriction, and YouTube has not disclosed whether its tools block political-figure prompts. None of the five platforms restricts political figures themselves from using generative AI tools within their own accounts. This gap represents a meaningful blind spot, given that every major platform has substantially expanded its in-house generation tools since 2024.

AI labeling may depend on platforms' own tools or self-disclosure. Meta, YouTube and TikTok apply automatic labeling to content made with their own tools, backed by C2PA metadata detection for content made elsewhere. Snapchat applies a visible watermark only to content generated through its native text-to-image features, with no confirmed evidence that it detects C2PA metadata or labels realistic AI content made elsewhere and uploaded. X detects C2PA metadata or flags undisclosed AI content, and users have no obligation to disclose AI-generated content that otherwise complies with the platform's rules. As a result, both platforms rely substantially on self-disclosure, which is likely to be insufficient when the content is designed to deceive.

Political advertising rules on AI-generated content are inconsistent and leave gaps. Platforms take substantially different approaches to disclosing or restricting AI-generated content in political advertising, and the resulting patchwork has left meaningful gaps heading into the 2026 midterms. Meta and YouTube (Google) require advertisers to disclose AI-generated or synthetically altered content in political ads, while X's dedicated Political Content ads policy does not mention AI or synthetic content. TikTok's ban on political advertising has not prevented political advertising from running on the app. Without a searchable US ad library, researchers have limited ability to independently verify whether such ads contain AI-generated content. Snapchat's political advertising policy also does not reference AI-generated content. Despite heightened public concern in 2026 about the risks synthetic media poses to elections, the absence of a cohesive, cross-platform standard leaves voters unable to predict how a platform will handle AI-generated content before it reaches them.

Restricted data access limits researchers' ability to assess the systems now shaping election information. Independent research is one of the few ways to check whether platforms are enforcing their own election policies, yet access to platform data remains limited, costly and uneven. Since 2024, these gaps have widened as platforms introduce new systems that come with little or no research access. No platform other than X shares its crowd-sourced fact-checking data. The chatbots

systematically, and no platform discloses how its AI moderation systems perform. As a result, the tools with the greatest influence over what content voters will see remain outside the scope of independent researchers, and harms may only come to light after the election is over.

Election content moderation increasingly relies on crowd-sourced and automated systems whose speed, accuracy and oversight remain difficult to assess. Research on crowd-sourced systems shows they often struggle during high-volume, fast-moving events, such as elections. Meanwhile, limited public information about the accuracy of automated tools and the extent of human review makes it difficult to assess whether platforms can effectively identify and respond to content that could threaten public safety or the integrity of the electoral process in November. This lack of transparency also limits the ability of users and independent researchers to identify potential bias in moderation decisions or assess whether permitted content is being over-moderated.

Platforms have continued to cut trust and safety staff since 2024, while disclosing little about the remaining internal resources dedicated to safeguarding elections. Major platforms have continued to undergo workforce reductions and restructuring, in many cases affecting trust and safety teams and increasing reliance on automated moderation. At the same time, platforms provide limited public information about staffing levels, language coverage or the extent of remaining human oversight in election integrity efforts. This lack of transparency makes it difficult to determine whether sufficient human capacity and expertise remain to address election-related risks.[1]

[1] ISD contacted the platforms examined in this report to share our assessment of their relevant policies, based on publicly available information, and invited them to provide corrections, clarifications or additional context. None of the platforms responded to our requests for comment.

In-Depth Analysis

AI chatbots are now standard across every platform, with limited transparency around election-specific safeguards

During the 2024 election, Meta and Snapchat were the only platforms in this review with widely available, integrated generative AI chatbots. As of 2026, AI assistants have expanded across all platforms examined in this report. xAI debuted Grok on X in 2023, and the feature became available at no cost to all X users in December 2024. TikTok began testing a chatbot named Tako in a limited rollout in the Philippines in May 2023; the feature remained in restricted testing for several years before TikTok widely rolled it out in February 2026. YouTube's comparable feature, Ask YouTube, was announced at Google I/O in May 2026 and initially made available to US Premium subscribers aged 18+, before expanding to general availability in August.

The rapid expansion of these assistants has, in some cases, continued close to the midterms. Meta, for example, launched Muse, a personal AI assistant powered by Muse Spark, in September 2026. Research published shortly after its launch raised concerns about the effectiveness of its safeguards for handling sensitive information. The limited time between the introduction of some of these tools and the election raises questions about how much opportunity there has been for independent scrutiny of their performance and safeguards.

While each platform has an embedded, conversational chatbot heading into the 2026 midterms, there is limited public information about how these tools will provide reliable election information. TikTok worked with Democracy Works on its 2024 Election Center but has not stated whether that partnership extends to its chatbot. Meta has stated that Meta AI will answer voting questions with local information or links to government sources, and in July said it engages with associations of election officials, without detailing the scope, frequency or impact. Google, YouTube's parent company, has announced a Democracy Works partnership for Search and Gemini, but not for Ask YouTube. The other platforms have not made comparable statements or documented similar action steps. In August 2024, however, X responded to pressure from state officials by having Grok direct election-related queries to Vote.gov—a reactive fix rather than a proactive partnership. ISD found no public documentation that this mechanism remains in place for the 2026 midterms.

Every major platform has expanded its in-house AI generation tools since 2024, but protections for political figures remain inconsistent

In 2024, most platforms offered only basic AI image generation tools, largely limited to still images generated from text prompts. Going into the 2026 midterm elections, every platform has expanded into more advanced generation, reflecting a broader industry shift from simple image creation toward full multimodal capability.

Meta and Snapchat have said they restrict their image tools from generating political figures. In 2024, Meta reported that its Imagine tool rejected 590,000 requests to generate images of Donald Trump, Kamala Harris, JD Vance, Tim Walz and Joe Biden in the month before the election. In April 2024, Snapchat stated that it had adopted "system-level restrictions on the generation of risky content categories, including the likeness of known political figures."

Restrictions like these are likely to protect widely recognized national figures better than local candidates, local election officials and civil servants,

who may be more vulnerable to being depicted by AI tools. The launch of Meta's Muse Image highlighted this exposure: in July 2026, Meta AI let users generate images of any adult with a public Instagram account by tagging them, without notifying the person. Meta removed the feature days later, stating that Muse Image blocks "violent, sexual, or defamatory imagery of real people." ISD found no public statement on whether it blocks political-figure prompts.

TikTok applies a similar block, but only within one of its own tools: Symphony Creative Studio instructs advertisers to "not reference celebrities," warning that "prompts that include the name or likeness of public figures are not supported and will be rejected" during video generation. ISD found no equivalent restriction for TikTok's other generation tools, whose output is instead governed by its content moderation policies. TikTok's Community Guidelines state that the platform does not allow AI-generated content that "misleads about matters of public importance or that harms individuals." By contrast, xAI's Acceptable Use Policy, which governs Grok Imagine, contains no provision specifically addressing political figures, candidates or elections as a protected category. Its relevant restrictions instead target non-consensual intimate imagery and general impersonation, making them narrower in scope than Meta's or Snapchat's stated restrictions. YouTube has not made an explicit statement on whether its in-house generation tools block prompts involving political figures at the point of generation.

None of the five platforms has published a policy on how public figures or candidates may use its integrated AI tools. The only restrictions identified apply to political advertisers, not to political figures' or candidates' own accounts. Meta and YouTube (Google Ads policy), for example, require advertisers to disclose AI-generated content in political ads, but those rules govern the ad itself, not who may use the tool.

Platforms lack dedicated policies for AI-generated political content

Several platforms have standing policies addressing AI-generated or manipulated political content, regardless of which tool was used to create it. YouTube requires creators to disclose realistic AI content that "makes a real person appear to say or do something they didn't do." TikTok's Community Guidelines prohibit AI-generated content that "misleads about matters of public importance or that harms individuals." Snapchat's Community Guidelines prohibit "spreading [of] false information that causes harm or is malicious, such as [...] undermining the integrity of civic processes, or manipulating content for false or misleading purposes (whether through generative AI or through deceptive editing)." Meta's organic-content labeling system (the 'AI Info' tag) applies broadly to detected AI content but does not specifically address political figures, unlike its advertising policy. This suggests Meta's most developed safeguards for AI-generated political content remain concentrated in its advertising products rather than organic posts. While consistent labeling is important for media transparency, current research suggests that it only partly limits the influence of AI-generated visuals.

Only Meta and YouTube have built AI-specific disclosure rules into their political ad products

Platforms take substantially different approaches to AI-generated content in political advertising. Meta and YouTube have introduced ad-specific requirements for advertisers to disclose or label synthetic or AI-generated content. X, by contrast, applies its general Synthetic and Manipulated Media policy to political ads rather than imposing additional AI-specific requirements on advertisements. TikTok prohibits political advertising in any form, although ISD research found that political ads have nevertheless appeared on the platform. Snapchat's political advertising policy also does not reference AI-generated content, although the platform states that human review of political ads checks for misleading AI use. Its April 2026 guidance states that AI-generated content may not be eligible for recommendation on

Spotlight but does not clarify how AI-generated political ads are treated under its advertising policies.

AI labeling may depend on platforms' own tools or user self-disclosure

All five platforms label content created with their own embedded generation tools, though the underlying mechanisms differ in depth. Meta labels outputs from its Imagine feature, stating directly: "We already add 'Imagined with AI' to photorealistic images created using our Meta AI feature." TikTok applies the most thorough system of the five platforms: content made with its AI Alive tool receives both an on-screen 'AI-generated' label and embedded C2PA metadata, which TikTok says persists "even if it's downloaded and shared off platform." YouTube extends comparable treatment to its newest tools, explicitly naming Nano Banana and Gemini Omni as features covered by its AI-content terms. Snapchat adds a visible watermark to images generated through its text-to-image features, describing it as a "clear, visual indicator that content is AI-generated." X applies 'Made with AI' labels to AI-generated content under its Synthetic and Manipulated Media policy, including content generated with Grok Imagine.

Meta, YouTube and TikTok have built robust systems, combining automatic detection with creator disclosure requirements. Meta reads C2PA Content Credentials to detect and label AI content made with other companies' tools. YouTube may automatically apply a label when content contains C2PA metadata, and creators cannot adjust labels on content made with YouTube's AI tools, content containing C2PA metadata or content labeled after manual review. TikTok's AI Alive tool applies both a visible on-screen label and embedded C2PA metadata that "persists even if it's downloaded and shared off platform." Snapchat's confirmed scope is narrower. ISD found no evidence that the platform detects C2PA metadata or labels AI-generated content created using external tools.

X relies most heavily on self-disclosure and shows the least evidence of independent detection. There is no confirmation that X automatically detects undisclosed AI content or C2PA metadata. Thus, outside of content generated directly through Grok Imagine, its labeling functions closer to an honor system.

YouTube's policy requires disclosure specifically when AI content is realistic enough that viewers could plausibly mistake it for authentic footage. Non-realistic content (such as a person riding a unicorn or a green-screen space scene) and minor edits (beauty filters, color correction or script assistance) are explicitly exempted. None of the five platforms requires additional or heightened labeling specifically for political content. YouTube's own listed examples requiring disclosure include "depicting a public figure stealing something they did not steal," but its policy does not establish a separate category or higher standard for political content. TikTok's Community Guidelines apply a broader standard than a label requirement, as they prohibit AI-generated or significantly edited content that "misleads about matters of public importance or that harms individuals."

Platforms are shifting from third-party fact-checking toward crowd-sourced moderation

Since 2024, platforms have moved toward crowd-sourced moderation similar to X's Community Notes model to complement or replace their fact-checking programs. Meta, specifically, ended its third-party fact-checking partnership in the US in 2025, launching its Community Notes-style model in April 2025. YouTube has piloted a similar feature since mid-2024, and TikTok launched its own version, Footnotes, in 2025, as a supplement rather than a replacement for fact-checker partnerships. For Meta and X, community-based moderation is the primary, and in some cases the only, mechanism for addressing misleading claims in organic content, including premature victory claims and allegations of election fraud.

While crowd-sourced models can foster trust and encourage diverse perspectives, research suggests

they face particular difficulties in politically charged, fast-moving environments. ISD's previous research on X's Community Notes—the longest-running crowd-sourced model of its kind—found that well-sourced, accurate notes on politically charged claims frequently failed to reach the 'helpful' threshold and were significantly delayed or simply never published. A manual review of the 25 most-voted Notes tied to the 2024 election found that, of the 11 Notes deemed accurate, necessary and consistent with X's standards, only two were published. ISD case studies of other fast-moving events found that only 10–20 percent of high-engagement false claims received a visible Note, with a median delay of 46 hours before publication.

These systems are also becoming increasingly automated. In July 2025, X announced plans to use AI to help write Community Notes, while subsequent changes introduced AI-generated 'Collaborative Notes' as an experimental feature still subject to community rating and approval. In May 2026, a study found that just eight AI accounts wrote half of all visible Community Notes on X, raising further questions about the capacity of the model to scale effectively while maintaining quality and oversight.

There is also a significant transparency gap. Meta, TikTok and YouTube do not provide public access to their contributor and Notes data, leaving outside researchers unable to independently assess how their systems are performing during the electoral window or whether they have addressed the shortcomings observed in X's model.

AI-powered moderation is expanding faster than transparency around its use

Apart from relying on crowd-sourced moderation, platforms are increasingly turning to AI to support content moderation at scale. Meta has said it uses large language models (LLMs) to provide a "second opinion" on some content before enforcement. In June 2026, the Financial Times reported that Meta had already replaced around 50 percent of human review requests with LLMs, with plans to reduce human review further for certain types of content. Meta said its initial tests showed that LLMs made 10 percent fewer mistakes than humans when enforcing policies against violating content and identified 10 percent more actual violations. TikTok states that it uses LLMs to "scale and improve content moderation." YouTube has not publicly described its moderation systems specifically as LLM-based, but says generative AI is used to improve the AI classifiers that identify content in violation of its policies.

The growing role of AI raises questions about how these systems perform when applied to fast-moving, politically sensitive content. Platforms provide limited information about which moderation decisions are made or supported by AI, including their error rates, the scope of their use, or the extent of human review required. Greater transparency would help assess how reliably AI-powered moderation can identify and respond to content violations, while providing greater visibility into the potential for errors that could result in legitimate political speech being incorrectly removed, restricted or otherwise limited.

Conclusion

The 2026 midterms will be the first US election conducted within this transformed, AI-powered information environment. Platforms have embedded chatbots, in-house image and video generators and moderation pipelines increasingly staffed by language models rather than people. ISD finds that while platforms have not ignored this shift, their response has been uneven, poorly documented and, in several cases, poorly suited to addressing the harms it is meant to prevent.

Three distinct patterns run through this report. First, generation has outpaced governance: platforms have introduced new AI tools—Gemini Omni, Nano Banana, Muse Image, AI Alive—within a matter of months, while policies governing their use in content involving political figures, candidates or elections remain underdeveloped or, in some cases, do not exist. Second, oversight has shifted from people to systems whose performance under election conditions remains largely unproven: third-party fact-checking has given way to crowd-sourced and AI-assisted notes, which have a documented record of missing time-sensitive corrections, while trust and safety teams have continued to shrink as platforms rely more heavily on automated moderation. Third, and perhaps most consequential, platforms have restricted the access researchers would need to understand the scope of these issues. Chatbot accuracy, moderation error rates, the scope of human review and the extent of AI-generated political content remain locked inside the platforms themselves, disclosed selectively, if at all, and typically only after significant delay. These gaps compound one another, creating an information environment in which the systems with significant influence over what voters see are also the ones about which the least is publicly known.

As AI becomes further embedded in the information environment ahead of the November 2026 as they emerge will be critical. The experiences of past election cycles suggest that such risks rarely remain hypothetical, yet the lack of transparency leaves little indication of whether or how platforms are prepared to respond.

ISD | Institute for Strategic Dialogue

Amman | Berlin | London | Paris | Toronto | Washington DC

Copyright © Institute for Strategic Dialogue (2026). ISD-US is a non-profit corporation with 501(c)(3) status registered in the District of Columbia with tax identification number 27-1282489. Details of the Board of Directors can be found at www.isdglobal.org/isd-board. All Rights Reserved.

www.isdglobal.org