A Platform Accountability Framework for Elections in the Global Majority

TECHGLOBAL INSTITUTE

A Platform Accountability Assessment Framework for Global Majority Elections

Authors

Diane Chang
Mark Schneider
Sabhanaz Rashid Diya

Contributors

Muthuri Kathure
Manisha Biswas
Jose Diaz Azcunaga
Fahad Ibn Siddique

TECHGLOBAL INSTITUTE

techglobalinstitute.com

Acknowledgments

We would like to thank the civil society organizations who provided critical input and feedback, and the partners who helped us distill learnings during the Trust & Safety Workshop at the Berkman Klein Center for Internet & Society, Asia Pacific Internet Governance Forum and Digital Rights Asia Pacific, which helped sharpen this report. Thanks to TrustCon and Mozilla Foundation for partnering with us to share the learnings from this report, and learn from a broad range of stakeholders about the challenges of advancing platform accountability in different contexts. We are grateful to Abigail Adu-Daako, Fadzai Madzingira, Muthuri Kathure and Lisa Garcia for leading these discussions. Thanks to Odanga Madung, Josh Lawson and Tim Harper for their feedback and contributions.

Overview

Elections represent one of the most critical civic moments for a country. By the end of 2024, over 70 countries will have held elections, making it the biggest election year in history and a pivotal moment in determining how the world organizes itself around democracy and geopolitics. People will have gone to the polls in countries with a combined population of over 4 billion people[1], of whom 80 percent are outside the United States (U.S.) and Western Europe.

In recent years, social Internet platforms[2] (also referred to as Internet platforms) have become central to the discussion on democracy. In particular, there is growing debate about their role in influencing political participation, polarization, social conflict, and electoral outcomes[3]. Social media companies have been associated with a range of harms targeting individuals (e.g., harassment of public figures and election officials), groups of people (e.g., incitement of group violence), or society at large (e.g., election misinformation that depresses turnout and has the potential to change election outcomes).

Much of the debate has focused on the U.S., especially in the aftermath of the presidential election in 2016[4]. A relatively smaller portion of evidence-based research explores Internet platforms' role in non-U.S. democracies. Existing evidence indicates that the majority of the world's social Internet users live outside of North America and Western Europe: more than 90 percent of Facebook's 3.9 billion active monthly users[5] live outside of the United States and Canada[6]. Despite these staggering numbers, only 13 percent of Meta's content moderation resources[7] are directed outside the U.S., which provides evidence of significant levels of inequity in researching, preparing for, and responding to civic events.

To address potential gaps in policies and enforcement, Civil Society Organizations (CSOs) across the Global Majority (also known as the Global South) play a crucial role in documenting, auditing and informing the public about harms on Internet platforms including during elections. However, these efforts tend to be fragmented[8], addressing siloed categories of harm, and the response from platforms appears piecemeal and infrequent. There are few systematic and holistic frameworks available to CSOs to assess platform response and, subsequently, hold them accountable.

This report aims to fill the gap by:

1. Outlining common harms occurring on Internet platforms (such as harassment and misinformation) that take place during elections;

2. Providing an overview of companies' rules to address these problems;

3. Explaining how Internet platforms detect and enforce against these problems; and

4. Providing a framework for how CSOs can:
   a. assess whether standard platform responses are adequate in addressing risks in their contexts.
   b. engage constructively in mitigating these risks ("Election Accountability Assessment Framework").

An Election Accountability Assessment Framework (EAAF) designed with Global Majority communities in mind is critical for several reasons. First, CSOs across the Global Majority have struggled to engage with Internet platforms constructively, owing to limitations in personnel access[9], limited research and data, and asymmetries in knowledge and access to resources. Second, existing CSO-led documentation on technology-mediated civic harms lack consistent methodologies that are accessible to the public and usable across geographies, cultures and contexts. Third, under-investment in company resources and a smaller number of academic research on platform accountability in Global Majority contexts limits the ability of CSOs to establish evidence-driven, universal advocacy strategies.

The EAAF hopes to address the aforementioned challenges by arming Global Majority communities with a structured playbook to assess Internet platforms' preparedness and performance during civic events—including elections—while identifying disparities among different platforms in different countries. In sum, it provides a contextually-sensitive baseline to evaluate, compare and advocate for more accountability and transparency from Internet platforms operating in Global Majority countries.

To illustrate how these current systems may contain disparities among different countries, especially in the Global Majority, we supplement analysis of common harms with case studies on how these policies, products, and processes apply to multiple Global Majority countries. We focus on countries that have already had or will have a general election in 2024.

A Few Important Caveats

Over the past decade, CSOs have been at the forefront of auditing, mitigating and engaging communities in the Global Majority on platform-mediated harms. Due to project-based funding and limited network access, CSOs are often constrained to monitoring social media platforms during critical civic events, such as election, war, or periods of democratic or political transition. This approach to resource allocation mirrors the strategy employed by Internet platforms themselves, where internal resources are directed to specific countries or communities based on immediate priorities. This cyclical focus or tendency to "parachute in" inadvertently undermines opportunities for a deeper understanding of platforms' impact on communities and is lacking in contextualization of platform-mediated harms within longstanding societal, political and structural trends. It is therefore essential for funders and interest groups to recognise the need for longer-term investment to investigate platform design, policies and behaviors.

The field of platform accountability frequently focuses on comparative analyses of the disparities between the Global Majority and Minority. This

approach has gained considerable traction since the leaked Facebook Files generated global media attention, exposing how companies adopt differentiated prioritisation matrices to decide which policies, products and processes to implement in a given country. However, this method inadvertently reinforced technology determinism, restricting interest groups, including CSOs, to the same frameworks imposed by the platforms they seek to hold accountable. In this report, we tried to demonstrate that even when an Internet platform implemented the same policies or products in the Global Majority as in wealthier countries in the Global Minority, it failed to address systemic issues and harms. The first principles guiding platform design, business models and decision-making are predominantly based on a specific set of normative values that differ significantly from those practiced in different regions worldwide.

Lastly, governments and civil society organizations (CSOs) in the Global Majority often call for the localization of policies or products or increased resourcing for their countries. Platforms, however, push back against these demands, citing the difficult trade-off between scale and consistency and emphasizing their reliance on a universal set of policies to minimize disparities. In less democratic countries, excessive localization risks enabling state-sanctioned censorship and repression. We urge researchers and practitioners to rigorously test this hypothesis, as scale is often driven by cost rather than consistency. This should not prevent platforms from contextualizing their offerings and developing products that address the specific needs of their user bases.

Pathways for harm on Internet platforms during an election

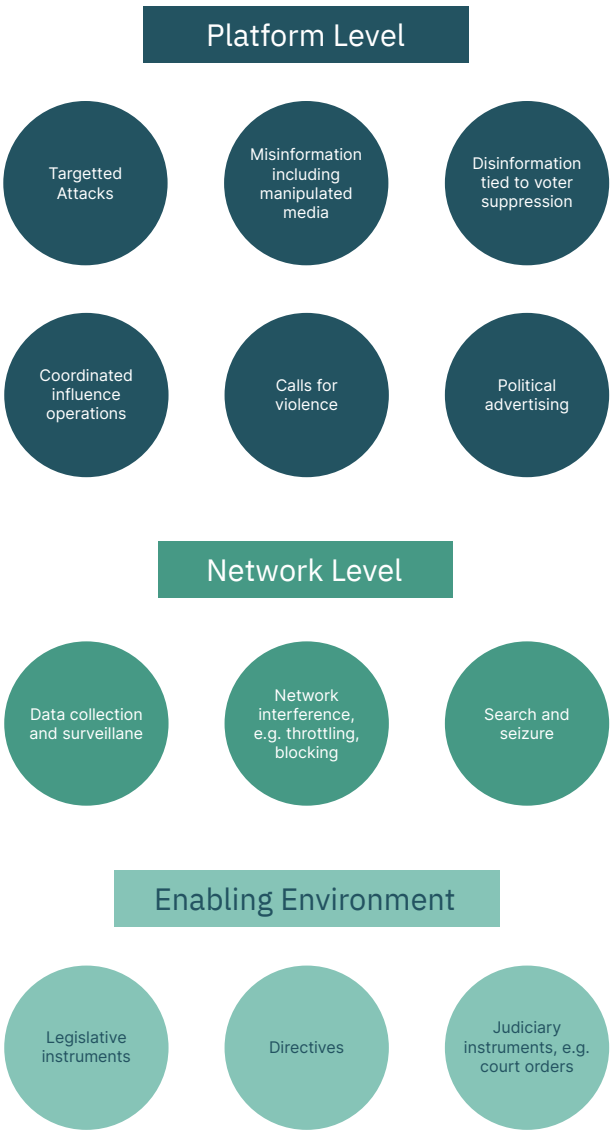
Much of daily communication for people around the world is mediated through social media platforms. For this reason, the social media ecosystem can be associated with several common risks leading up to, during, and immediately after an election. We categorize them broadly as platform, infrastructure and external pressure risks, as defined below.

- Platform risks: Risks or harms to the online information ecosystem that affect how people express and organize themselves. Common examples include misinformation, disinformation, and calls for violence. Platform risks can occur at two levels:
  (a) Content (what people create and share) and
  (b) behavior (actions associated with individual or multiple accounts).

  People become exposed to these risks through different mediums—public, private, and encrypted—that warrant varying degrees of policies and response.
  (a) Public mediums include commonly used platforms such as Facebook, YouTube and TikTok, where users can view and interact with content and accounts, and where viewing is managed through personalized settings.
  (b) Private mediums, such as Facebook Groups, facilitate more closed interactions between groups of individuals that can't be accessed

by the general public without an invitation by existing members. (c) Encrypted mediums, such as Whatsapp and Signal, represent a more private, end-to-end encrypted interaction between two or more individuals.

Pathways for Online Harm During Elections



- **Infrastructure risks:** Risks or harms that a telecommunication network is exposed to, typically affecting device access or the connectivity infrastructure. Common examples include network interferences, also known as Internet shutdowns, that involve blocking and/or throttling of Internet access points.

- **External pressure risks:** Risks that manifest as a result of the legal and policy environment, resulting in negative downstream impact on how people access and share information online. External pressure risks can also take the form of threats by governments to detain or imprison company personnel, as a means of influencing the companies' decisions.

In this report, we focus primarily on platform risks. In limited cases, we provide examples where infrastructure or external pressure risks have exacerbated platform risks, resulting in an adverse impact on democracy and civic participation. Throughout the report, there are individual and societal harms that can be evaluated based on prevalence and severity.

How Internet platforms determine context

Major Internet platforms rely on several external and platform-specific signals to determine how they should respond to platform risks. A common example of an external signal can be regulatory obligations, while platform-specific signals can include number of users and staffing. In many resource-constrained Global Majority countries, it may be challenging for platforms to gather

sufficient external data to definitively assess the conditions leading up to an election. Among others, external signals indicating whether a country is democratic or not are of particular significance. Many research and public indices (for example, Freedom House's Freedom on the Net and Economist Intelligence Unit's Democracy Index) measure how each country is faring on democracy metrics and assign them a score to determine if a country is 'free' or 'fully democratic', 'partly free' or a hybrid democracy, or autocratic. A 'fully democratic' country will have a consistent record of fair and transparent electoral processes; separation of administrative, judiciary and electoral bodies; strong legal and political safeguards around freedom of expression and political participation, and so on.

If a country is deemed to be democratic, companies are more likely to introduce levers to mitigate platform risks aimed towards safeguarding candidate and voter participation in elections. The standardized menu of election integrity levers is designed for ideal democratic conditions. These levers are scaled down if an election is deemed less transparent and non-participatory. This is based on the assumption that the interventions will have limited impact on users' communication, underperform or be abused by bad actors.

In this report, we focus on the implementation of the standardized menu of integrity levers, however, we fully acknowledge that the signals and menu themselves are a limitation in responding to context-specific risks. We hope to dedicate future research on the latter.

How Internet platforms protect platform integrity

Major internet platforms' efforts to protect the integrity of elections fall within broader ongoing efforts of these platforms to intervene when users' safety and trust are at risk. We adopt the "ABCD Disinformation Framework" [10] to describe the ABCD's of Platform Integrity Levers (i.e. changes to how features and systems work) that Internet platforms deploy to reduce users' exposure to a broad range of online harms, including in the context of elections. This

ABCD's of Internet Platform Integrity Levers	
Actors	• **Taking down & limiting** malicious actors, such as temporarily restricting access to accounts or features. Can result in removal of accounts for certain violations or other penalties.
Behaviors	• **Taking down** accounts, groups, pages, and networks that engage in prohibited behaviors such as misrepresenting who they are, coordinating harm, spreading coordinated information operations. May be **proactive & automated** or result from manual **investigation**.
Content	• **Removing or restricting** of content that violates Community Standards or Community Guidelines
Distribution	• **Reducing visibility** of harmful content that is assessed to be harmful, but not violating policies. In some cases, this is coupled with **labeling** and/or **amplifying** of information from authoritative sources and fact-checkers. • **Restricting the reach** of rule-breaking groups, and group members, such as deboosting or removing them from recommendations. • **Provide incentives through monetization or creator programs** to promote high quality content.

Proactive vs. Reactive Approaches to Content Moderation	
Proactive	Companies proactively enforce content rules on their platform using automatic and human detection (e.g., automatically removing slurs when detected).
Reactive	Companies pursue reactive enforcement through escalations of problematic content surfaces internally or often by CSOs. In escalations, internal and external stakeholders weigh in on the appropriate action for potentially problematic content. Meta's Oversight Board is a publicly observable example of this reactive approach.

framework, while the same as that used by Internet platforms, warrants scrutiny from researchers and interest groups to determine whether it adequately addresses the risk vectors prevalent in Global Majority contexts.

Overview of election integrity on internet platforms

To illustrate the role that Internet platforms play in countries' elections, this report focuses on four major platform risks during elections under the integrity framework: Election integrity and voter interference; misinformation; foreign and domestic interference through coordinated and/or inauthentic information operations; and account security. This is not intended to be an exhaustive set of all common online risks during elections, but rather a selection made to illustrate a range of possible intervention levers. We want to reiterate that operating within the constraints to these risks reinforces technology or platform determinism and future research should explore the taxonomy of platform-mediated harms applicable

	Policy	Process	Product
Voter Interference & Election Integrity	Voter suppression (unpaid content), manipulated media policies: Violating content is removed. Voter suppression and election delegitimization in ads is banned.		Election Information Panel, Voting Information Center, Voting Reminders: Elevate relevant information on elections and how to vote in prominent product surfaces.
Misinformation		Fact-checking program: Independent fact-check partners investigate, debunk, and rate narratives or stories receiving wide circulation.	Fact-check labels: Show users fact-check rating and debunked information provided by fact-checkers, displayed on relevant content.
Foreign & domestic coordinated deceptive behavior	Coordinated inauthentic behavior, deceptive practices, coordinating harmful policies can result in take downs of networks and entities.	Disclose: Takedowns may be publicly disclosed by platforms after action is taken or in regularly published disclosures.	Transparency labels: Accounts by state-controlled or state-backed media outlets are labeled as such, in order to provide users with transparency.
Account Security (of candidates, journalists, activists)		Partnerships: On-going engagement, training and, feedback gathered through external engagement relationships or trusted partner programs.	

to diverse Global Majority contexts. The below chart shows an example of some policies, processes, and products that may be applied by platforms to intervene on these risks.

Afterwards, we elaborate in further detail on some of the policies, processes, and products that Meta (which owns Facebook, Messenger, Instagram, and WhatsApp), Google (who owns YouTube), and ByteDance (who owns TikTok) may have in place to address these risks.

US Midterms 2022[111213]

We take the U.S. 2022 midterms as an example to showcase the range of policies, products, and programs that companies leverage to prepare for and respond to an election. While we cannot definitively conclude using publicly available materials whether platforms had an outsized investment in preparing for and responding to the U.S. midterm elections, we cite several sources that allude to a wide and diverse range of interventions that were introduced at the time. We use this example to illustrate a high investment election and evaluate whether companies have applied similar interventions to Global Majority countries' elections. It is crucial to note, however, that this report does not advocate for the same interventions as the U.S. to be launched elsewhere in the world. Instead, it evaluates the extent to which companies research and launch interventions for the unique risks in Global Majority countries.

Meta announced a fairly extensive set of efforts for the U.S. midterm elections around three months before election day, included that they had 40 teams working on the election, and had added support in both English and Spanish languages, which was an improvement made based on feedback from civil society organizations during the 2020 election[14]. YouTube announced around two months before election day that they were prominently recommending content coming from authoritative national and local news sources like PBS NewsHour, The Wall Street Journal, and Univision, the latter being a similar recognition of the need to support a growing minority voter population[15]. TikTok announced around three months ahead of election day that they planned to add labels to content identified as being related to the midterm elections as well as content belonging to politicians, political parties and the government, to provide transparency to users[16]. Twitter (now X) activated their civic integrity policy 3 months prior to the midterms, with redesigned labels for misinformation, pre-bunks, state-specific event hubs, candidate labels, better recommendations to filter misleading tweets, and a dedicated explore tab for national news and voter education announcements. To note, we did not find comparable and consistent levels of transparency from Internet platforms about other elections around the world.

Philippines Election 2022 and Differences among Global Majority Elections

In comparison, a country in the Global Majority that held an election in 2022 was the Philippines, home to nearly 90 million social media users[17]. The Southeast Asian country has topped the global list on Internet use with people spending an average of ten hours of screen time every day[18]. Since 2017,

YouTube launched several election features [during the Philippines election], including candidate panels for the president and vice president, as well as panels to inform users on how to vote.

after the country's prior general election of 2016, multiple reports warned of the use of social media as a tool for "digital repression" and amplifying disinformation[19]. The situation on the ground significantly worsened under the Duterte administration's Anti-Terrorism Act of 2020. State officials could label and brand individuals as left-wing, communist, or terrorist, also known as 'red-tagging', which would result in arrest and detention without a warrant[20]. The legislation was widely abused in the Philippines to persecute critics of the state, journalists, and political dissidents through the amplification of false allegations using social media[21], thereby exposing targets to heightened risks of unwarranted arrest, harassment and physical violence. Additionally, historical revisionism of the brutality under former President Ferdinand Marcos– father of 2022 presidential candidate and eventual victor Ferdinand "Bongbong" Marcos, Jr.– through white-washing and nostalgic portrayals of the period, was reportedly prevalent on social media for several years leading up to the election in 2022[22].

TikTok[23] and Meta[24] made public announcements on how they were preparing for this election one to two months before the May 9th election day. The two companies announced their partnerships with the Commission on Elections (COMELEC) to direct users to a dedicated election microsite. Meta did not provide information on the number of teams focused on these elections or other investments for combatting platform risks during this election. TikTok additionally partnered with the public affairs media organization GMA News and Public Affairs to launch a campaign to educate users about the elections, starting around two months before election day.[25] YouTube launched several election features, including candidate panels for the president and vice president, as well as panels to inform users on how to vote[26].

On reviewing the companies' public announcements referenced above, we observe that Internet platforms adopted varied responses in the Philippines, indicating there was likely inconsistent enforcement on harmful content across the different platforms. Human rights groups and researchers allege that the fragmented, often lagging, response from Internet platforms, and western-centric policies that inadequately incorporated local context, resulted in what they see as rampant disinformation that benefitted Marcos Jr.'s campaign[27]

These inequities are evident of systemic gaps in how internet platform policies are designed and implemented, which result in a disproportionate impact in Global Majority countries. While we delve into select harm areas below, at a high level, any gap analysis should fundamentally lead with three questions.

1. Where and how are integrity decisions made for a given country, especially during major civic events like elections?

   Leadership of the major Internet platform companies is heavily concentrated in Silicon Valley,[28] in addition to a smaller group of lobbyists and decision-makers in Washington, D.C. While there are senior staff at a regional level, they are not consistently empowered to make final decisions on sensitive content or product that could impact an election in the majority of cases[29]. Further, product and engineering teams assigned to support civic events are largely based in the U.S., resulting in significant lags and contextual knowledge gaps in developing, deploying and refining products aimed at Global Majority countries[30]. While decisions are never made unilaterally inside companies and typically involve multiple teams, the geographic distribution of relevant teams and decision-making structures can become impediments, especially in crisis scenarios. There is limited transparency around how decisions are made, and how quickly decisions can be made, on high-impact election-related matters outside of the standardized risk assessments, especially in Global Majority countries where resources are limited and risks of violence and unpredictable political conflict can be relatively high.

   In these contexts, decision-makers need to be well-informed of political and policy contexts in real-time. Moreover, research has shown that the concentration of power in Silicon Valley has resulted in structures and decisions that are guided by business interests and simple technical solutions,[31] and fail to acknowledge and adequately respond to varied sociocultural contexts in Global Majority countries.[32]

2. How well-resourced is the market for language-sensitive artificial intelligence (AI) moderation systems?

   Criticism of content moderation in Global Majority languages often revolves around the number of human reviewers (or lack thereof) deployed by the platforms[33]. This is important, but a narrow view of moderation risks overlooking the fact that it is neither practical nor commercially feasible to have *enough* human reviewers to meet the ever-increasing volume of content. To address the large volumes of harmful content, an Internet platform needs to consistently and comprehensively invest in language-sensitive automated moderation systems that use artificial intelligence to detect and classify harmful content (also known as classifiers), across a wide range of harms.

   Automated or AI moderation systems come with their own challenges, especially for low-resource languages– those that are widely spoken in the world, but represent a smaller share of text on the Internet. Because there is less digitized text available to train AI systems in low-resource languages, Internet platforms rely on a process called cross-lingual

Despite over 50 national elections slated for 2024, by the end of 2023, major internet platform companies had published limited information on how they were preparing for elections in 2024.

transfer[34]. This essentially transfers lessons of an English-trained system to low-resource languages by using machine-translated text[35]. When content that is language- and context-specific is moderated through the intermediary of English, it poses serious risks of missing out on hate speech, violence incitement, and harassment in Global Majority countries.

Internet platforms are not transparent about how their multilingual AI moderation systems work, how they have been scaled, which areas of harm they have been trained on and how effective they are at tackling harmful content when scaled globally.

3. How transparent is the platform about its approach to election integrity in the market?

Internet platforms publish information on their efforts for elections in different countries, but this is infrequent and applicable to a limited number of countries.

Transparency is one of many critical aspects in how CSOs and policymakers can understand how Internet platforms are preparing for an election. In 2023, when at least 49 national elections were conducted around the world,[36] platforms provided varying levels of transparency on their efforts for elections around the world. TikTok published details on their efforts for Nigeria's 2023 general elections on their public blog[37]; however no public information was readily available for any other elections that year.

Meta's newsroom contained information on its plans to protect elections for only three countries[38]. X (formerly known as Twitter) has not updated its public blog on country-specific elections since the fall of 2022[39], and very limited official information was publicly available on YouTube's approach to any elections in 2023.

Despite over 50 national elections slated for 2024, by the end of 2023, major internet platform companies had published limited information on how they were preparing for elections in 2024. Meta published an update stating that its plans for 2024 elections were largely consistent with efforts for past elections, with one notable change in its advertising policy to require advertisers to disclose if their political ads are created or altered using artificial intelligence[40]. The update linked to a fact sheet outlining how the company was preparing for the U.S. election[41], but did not reference any other countries' elections.

YouTube publicized its updated efforts specifically on election misinformation in preparation for the U.S. and E.U. 2024 elections[42][43], and Google announced a similar new advertising policy in the fall of 2023 to require disclosure on political ads that are altered or generated using artificial intelligence[44]. TikTok did not provide any information on its preparations for any 2024 elections in its newsroom.

By the spring of 2024, major internet platform companies published information on their preparations for around 10 elections in 2024, including in the U.S., E.U., India, Mexico, Indonesia, Brazil, Mexico, South Africa, and the U.K.. Some of the election plans and policies seem to apply globally, while a few of the specific plans and products seem to be localized to each country. To address the novel challenges posed by generative artificial intelligence in 2024, Meta, Google, and TikTok signed up for a symbolic voluntary agreement among technology companies to combat deceptive uses of artificial intelligence in all 2024 elections. Meta published elections plans for Brazil, India, the EU, the UK, Mexico, and South Africa. Notable new approaches include an advertising global policy update to require advertisers to disclose if their political ads are created or altered using artificial intelligence[45]. The launch of a fact-checking helpline on WhatsApp in India to debunk AI-generated misinformation[46], and trainings of civil society groups in South Africa on the safety of marginalized communities along with the publication of country-specific resources[47]. Google announced preparations of its products, including YouTube, for the U.S., E.U., and India 2024 elections[48][49][50]. These include a similar new advertising policy requiring disclosure on political ads that are altered or generated using artificial intelligence,[51] and a requirement that YouTube creators disclose content that is synthetically created or altered to look realistic[52]. TikTok announced election plans and policies in Indonesia, Mexico, South Africa, the EU, U.S., and U.K.[53][54][55][56]. It instituted a global ban of misleading "manipulated content," including AI-generated content of public figures endorsing a political view.[57] It shared that it had signed a memorandum of understanding with the Federal Electoral Tribunal of Mexico (TEPJF) to discourage disinformation and promote transparency and accountability. Tiktok also announced it had set up a dedicated channel for the Indonesian Election Supervisory Agency (Bawaslu RI) to report misinformation to TikTok, and would partner with Indonesian creators and experts to educate users on misinformation and responsible content creation. For South

For South Africa's election, TikTok announced a partnership with Code for Africa to ensure 'the accuracy of content in numerous South African languages' and touted the contribution of local creators to a media literacy campaign with content in languages including English, Afrikaans, isiZulu, isiXhosa, and sign language.

Africa's election, Tiktok announced a partnership with Code for Africa to ensure "the accuracy of content in numerous South African languages" and touts a the contribution of local creators to a media literacy campaign with content in languages including English, Afrikaans, isiZulu, isiXhosa, and sign language.

Appendix 1: Detailed company plans for the U.S. 2022 election

TikTok: For the 2022 U.S. Midterm election, TikTok launched a dedicated Election Center within the app[58], which provided information from reliable third-party sources in more than 45 languages, including English and Spanish. The Election Center directed users to information on how and where to vote, and what would be on their ballots. The company launched labels to accounts belonging to government entities, politicians, and political parties. Some of the labels as well as certain hashtags related to the election were linked to the election information guide. The company stated that, based on learnings from the 2020 elections, it planned to launch the Election Center six weeks earlier than in 2020.

During the 2022 elections, TikTok seemed to utilize its standard tactics to combat misinformation. The company partnered with fact-checking organizations to assess the accuracy of content in purportedly over 30 languages, working in tandem with their internal investigation and moderation team. When content was flagged as misleading by fact-checkers, TikTok applied standard treatments in the app such as informing users and prompting them to reconsider before sharing. Content that was being fact-checked or which couldn't be substantiated through fact checking but was considered potentially misleading, it was removed from recommendations. Consistent with their standard policy, political ads were not allowed during this election.

Meta: Similar to past U.S. elections, Meta provided a Voting Information Center in the Facebook app that contained information from authoritative sources. Facebook launched a new feature that showed links to official information on how, when, and where to vote, among search results when users search for terms related to the election.[59] Instagram launched "Register to Vote" and "I Registered to Vote" stickers in English and Spanish that users could add to their Instagram Stories. When users clicked on one of these stickers in a Story, they were directed to voting information from their state.

The company stated that it deployed "advanced security operations to fight foreign interference and domestic influence campaigns," and continuously reviewed content for violation of its policies, including those on election interference, hate speech, coordinating harm, and bullying and harassment. The external partners Meta worked with and that the company revealed included: State and local election officials, the federal Cybersecurity and Infrastructure Security Agency, industry peers, the National Association of Secretaries of State and the National Association of State Elections Directors. Meta applied warning labels to content that made false claims about the election results and removed content that provided false

information on when voting took place.

The company stated that it would reject ads "encouraging people not to vote or calling into question the legitimacy of the upcoming election." Similar to its policy in 2020, it did not allow new political, election, and social issue ads to be run during the final week of the election campaign.

YouTube: YouTube launched a number of features to elevate authoritative news and information leading up to the 2022 U.S. election. It displayed a panel of relevant election information, in English and Spanish, at the top of search results and below videos about the election. The company stated that it prominently recommended videos from authoritative news sources when users searched about the election, and also limited the spread of harmful election misinformation. YouTube displayed reminders on registering to vote and other election resources on its homepage before the election. On election day, YouTube displayed labels on search results and under videos about the election that linked to Google's dedicated feature showing real time updates on election results[60].

YouTube reiterated its standard enforcement processes, such as for content that incited violence or contained certain types of election misinformation. The company's September 2022 announcement on election plans stated that it had already been removing content violating its policies, including content that claimed the U.S. 2020 election was rigged or stolen.[61] The company also launched a campaign across YouTube to educate users on media literacy skills and techniques.

Key Platform Risks and Relevant Response

Election Integrity & Voter Interference

The problem

Several categories of online content can be weaponized to interfere with voter participation. Misleading or incorrect information on when and how to vote, how elections work, and election results, can all have the effect of deterring people from voting, disenfranchising voters, or causing mistrust in elections. Additionally, a broader range of misinformation, such as on politicized topics or claims about candidates, can have the effect of eroding trust and deterring participation. This section covers the first category of misleading information. *Please see the next section for more on misinformation.*

Misleading information related to the integrity of elections is harmful because it can erode trust in democratic processes and institutions, and even elevate the risk of offline violence, as has happened in the U.S., Nigeria, Brazil, and other cases . Beyond risks of mass violence and lost faith in the democratic process, misleading content can affect individuals involved in the election process, such as poll workers. For example, in the U.S. state of Georgia, poll workers accused of fraud in the 2020 election received numerous death threats[62]. This means that platforms must be cognizant of misleading content related to the voting process in general and the specific individuals involved in the voting process to the extent that this is a salient form of attack in a particular country.

How do companies identify and address voter and election interference?

The major internet companies have policies prohibiting content containing false information on election timing and voting procedures, or by encouraging or intimidating voters of particular groups to abstain from voting. Policies from YouTube, Facebook, Instagram, and TikTok state that content that can cause offline harm– including harm to election processes– are prohibited. Major companies either ban or explicitly correct content with incorrect election results, although this is limited to national elections to the best of our knowledge.

Platforms detect and remove specific, narrow types of misleading content that are outlined in their community guidelines and policies, such as content that could contribute to offline harm, content that misrepresents the date or place of an election, or content that intimidates voters. During some elections, civil society, government, political, and other external partners may send identified voter interference content to the internet companies, which is then reviewed on a case-by-case basis.

Major Internet platforms also provide users with authoritative and reliable information related to elections through

several types of election information products. One type of product is a dedicated "hub," such as a web page or whole screen, that contains a range of information on an election and how to vote. A second type of product is a panel or module, containing authoritative information on specific topics related to elections, that is inserted within the context of other content, such as between user-generated posts or above search results. The third type of product is a label, warning screen, or prompt that is displayed on top of user-generated content that doesn't directly violate the community guidelines but may be unreliable, in order to surface authoritative information and context to help users evaluate the underlying content more discerningly. An example is a label on a post that indicates that the post was created by a state-backed media outlet, or a label that directs users to an authoritative third-party source of information on how elections are administered. These labels might link to the information hub on the same platform, which generally cite information from reliable third-party sources, or link directly to an authoritative third party's website, such as the official election results page of an election commission.

Internet companies have a separate set of policies and solutions for some types of election misinformation that fall outside of the above categories, and can be debunked by fact-checkers. For more detail on this set of policies and solutions, see the "Misinformation" section.

Policy and enforcement gaps

It is unclear whether internet platforms' election integrity policies are in effect and enforced for all elections around the world, if enforcement is based on factors such as the type of election (e.g. national versus regional/local), or whether a given country is selected or prioritized. These election integrity and voter interference policies cover a very narrow set of topics that can undermine elections. However, many other types of content that can de-motivate voters and sow distrust in election processes, such as vague, questioning statements and speculation, are allowed on the platforms in lieu of escalations.

Gaps in Global Majority countries

While several of the major internet companies have provided a centralized destination on their platforms for authoritative information during past U.S. elections, the availability of these hubs or modules has been more limited in other countries. The development of these hubs may be limited by available resourcing to establish partnerships with authoritative information sources, as well as by challenges on what sources are considered authoritative in countries where election authorities may not be considered reliable, or trust in unbiased media institutions is low.

In fast-changing situations where election results may not yet be available or are not yet verified, and therefore there are no established facts yet for fact-checkers to rely on to debunk circulating rumors, platforms have to resort to other tactics to tackle content that may cause mistrust in the status

of election results. However, there is no public information on how platforms typically respond to such scenarios in different countries, including in countries where official results may be inaccurate or unreliable. This is relevant to the following specific gaps:

- Inaccurate claims about election results were corrected by some social media companies in the U.S.' 2020 and 2022 elections and in Brazil's 2022 elections, but it is not clear if this has been done in smaller markets among Global Majority countries.

- Tools such as TikTok's Election Center and Meta's Voting Information Center, containing authoritative information on elections such as voting times, may only exist in the U.S..

- Even if explicit claims of false results and false election times are corrected, content containing more indirect and implicit language with similar meaning is unlikely to be effectively detected by automatic systems.

- While escalation policies are a major tool for enforcement– where companies decide on enforcement on a case by case basis– this process is highly dependent on the availability and capacity of employees, and likely to be biased toward the most high profile cases that gain the attention of company staff.

Evaluating platforms' effectiveness

Platforms' ability to uphold election integrity and prevent voter interference in a given country is dependent on how they prioritize and enforce on content that makes false claims about election process, procedures or results. The following questions may guide in evaluation using the EAAF:

- Are platforms proactively and automatically enforcing election integrity and voter interference policies during your country's elections? To which elections (national, regional, local) do these policies apply?

- If a candidate in your country posts or shares content that violates an election integrity or voter interference policy, is that content removed?

- How effectively are platforms able to detect content that contains voter interference in the major languages in your country?

- How effectively are platforms able to identify context-specific, and coded language that de facto contains voter interference? How effectively are companies collaborating with organizations on the ground to elevate and interpret such content?

- How do platforms respond to incorrect claims about election results? How do platforms handle incorrect claims about the results of past elections?

- Do the platforms provide an authoritative source of information in

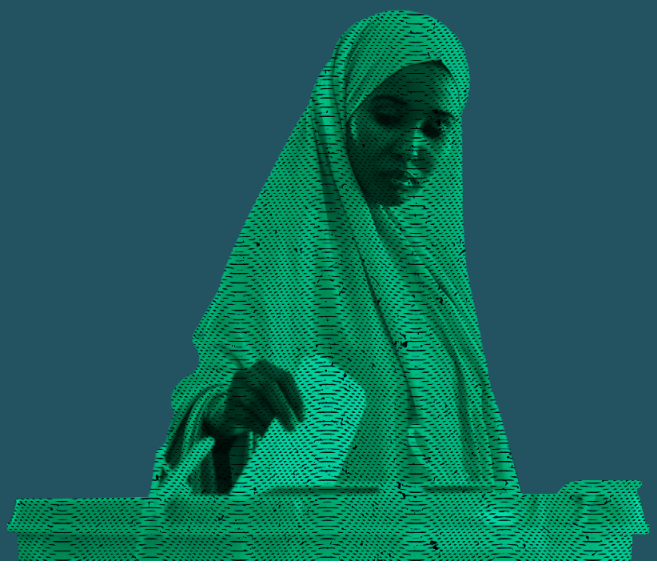
your country on election and voting rules and on election results? Do platforms prioritize reliable news sources with sound reporting and verification practices over low quality news sources?

- Do companies have a process and employees available to escalate content identified by civil society organizations that seems to undermine election integrity and voter participation?

Case Study: Election integrity in Nigeria

Nigeria held its presidential and national elections in 2023 with an estimated 93.5 million registered voters[102]. There are 30 million social media users, with WhatsApp as the leading platform used by 90 million Nigerians[103]. Research from local CSOs and academics indicated that social media plays a significant role in how Nigerians consume information that then determines their civic participation and decision to vote[104].

In the lead-up to the elections, people reportedly saw stories from unfamiliar media organizations on X and Facebook, often forwarded across WhatsApp, that contained false and misleading content to elevate or discredit candidates and sow confusion about voter participation[105]. According to a BBC investigation, these websites behind social media accounts were likely established during the time of elections, producing as many as 700 stories per month. One of the websites, *Parallel Facts*, conducted two live conversations with third-place presidential candidate Peter Obi on Twitter's Spaces feature for live audio conversations, after which *Parallel Facts*' engagement grew fourfold within two months, with up to 40,000 mentions by July 2023. *Parallel Facts* owner Kingsley Izuchukwu Okafor previously posted a photo of Peter Obi with the comment, "Obi is the man". *Parallel Facts* published stories and amplified on social media indicating Nigeria's Independent National Electoral Commission was "giving APC 25% of the votes" (All Progressives Congress, or APC, is one of Nigeria's two major political parties) despite lack of credible evidence confirming the allegation. *Reportera*, also established in the lead-up to the elections, published verifiably false stories, including alleging Bola Tinubu—who secured a narrow victory in the election—actually came third. Nigeria experienced a fiercely



contested election with second and third place presidential candidates alleging widespread voter fraud, which is now being contested in courts[106][107].

Ahead of the election, Meta announced that it planned to moderate content in Yoruba, Igbo, and Hausa, as well as remove out-of-context photos and videos that falsely depict ballot-stuffing, acts of violence and weapons[108]. Meta's publicly-posted Community Standards are only available in English and Hausa in Nigeria, the latter being one of the 3 major languages spoken in the country. A voter education chatbot was launched on WhatsApp in collaboration with a local non-profit[109], but the company did not mention having any additional features that would redirect users to reliable information when they search for terms related to the election. Further, despite well-documented evidence about Cambridge Analytica attempting to influence Nigeria's past elections[110], including dissuading voters from participating, Meta did not introduce any measures to reject or take measures against ads that could discourage voters or question the legitimacy of the election.

TikTok announced its in-app guide, known as an Election Hub, ahead of the Nigerian election[111]. In addition to utilizing its standard tactics to tackle voter fraud and misinformation, TikTok also launched labels for content identified as being related to the 2023 election. However, the company did not provide specific information on whether there were disclosure labels for accounts belonging to government entities, politicians and political parties.

YouTube did not publish any readily accessible information about how they planned to tackle risks to the Nigerian election.

Misinformation

The problem

Aside from the narrow types of misleading content about the mechanics and logistics of voting that are covered in the section above, other kinds of election-related misinformation can mislead people and sow distrust in elections. These other types of election misinformation are often fast-changing, highly contextual, and sometimes challenging to debunk. Common examples include claims or rumors about specific candidates, misleading content about typically nonpartisan issues that become suddenly politicized in the context of a specific election, or vague statements such as speculation about an election's reliability.

How do companies identify and combat misinformation?

Misinformation is treated differently from the prohibited categories of content outlined in community standards

and guidelines, because there is no way to establish a comprehensive list of what is prohibited, due to the contextual and fast-changing nature of misinformation. Instead, platforms act on these types of content according to misinformation policies and processes, in close partnership with third party fact-checkers. These fact-checking partners review and rate content as misinformation. Fact-checking partners are independent entities, but may receive some funding from the internet platforms to support their fact-checking work.

Platform companies and their fact checking partners identify potential misinformation "claims" in content on their platform through a variety of sources, such as automated detection of false information that matches content already debunked by fact-checkers, requests by government to take down content they believe to be misinformation, or reports by trusted external partners (such as CSOs). Their fact-checking partners also independently find, debunk, and rate false claims.

The platforms then attach display labels on top of the debunked content that show users the fact-checkers' rating and debunking information. Platforms may also reduce the circulation of the debunked content on their platforms so that fewer users are exposed to it, such as by demoting such content, or removing it from recommendations. They may also use artificial intelligence to detect new content that is likely to contain false information, and send the new content to fact-checkers. During some elections, civil society, government, political, and other external partners may sometimes send identified misinformation content to the companies, which is then reviewed on a case-by-case basis.

Policy and enforcement gaps

Common gaps across major social media companies are as follows:

- Rules are broad, which makes it possible to act on a wide range of concerning content, but not concrete enough for the public to have certainty on what is enforced against in practice.

- Processes for identifying and mitigating misinformation are generally opaque. This means that while the policy may be broad, it is unclear what content is actually reviewed by fact-checkers proactively, and what are criteria and techniques for them to identify content to review and debunk. It is also unclear what content is surfaced to companies by external parties for case-by-case review, what criteria determine which content is surfaced to third party fact-checkers to review and debunk, what sources of content issues get the most access to company insiders, and the extent to which risks to marginalized populations (e.g., religious minority demobilization efforts) are addressed as compared to risks faced by dominant groups on a platform within a given country.

- Platforms are hesitant to be "arbiters of truth," thus relying on third party fact-checkers and sources of

authoritative information. However, processes relying on human fact-checking are challenging to implement programmatically, across different countries, and at large scale. Human fact-checkers naturally have limited capacity, which means content that has less "reach"— such as misinformation targeting minority language populations, or misinformation at the regional or local level– is less likely to get fact-checked or removed from platforms unless it is identical to previously fact-checked content. Additionally, fact-checking is a reactive endeavor, which means users can be exposed to new misinformation claims before fact-checkers are able to review and debunk those posts.

- It is unclear if variation on fact-checked content using different language but capturing the same meaning is detected and enforced.

- Misinformation is inherently hard to define. Content that the public may consider "misinformation" due to its misleading nature, but which is difficult to debunk, generally will be permissible on the platforms. For instance, content that characterizes a brutal historical event in a misleading tone that minimizes the effects, but does not explicitly dispute the facts surrounding the event, is generally not considered eligible for fact-checking.

- Politicians or certain content that the companies consider "newsworthy" may not be eligible for removal or fact-checking, unless they fall under criteria for being considered "harmful." What is considered to be "newsworthy" can be highly subjective across countries and contexts[63].

Gaps in Global Majority countries

Platforms' capacity for fact-checking is dependent on the number of fact-checking partners they have available and their capacity relative to the volume of content they need to review and debunk. Companies may have fewer fact-checking partners in Global Majority countries due to the unavailability of fact-checkers, the company's decisions on how to prioritize resources, or due to external factors such as government persecution of journalists and news outlets, resulting in fewer organizations and individuals willing to participate in the program[64].

According to The Reporters' Lab at Duke University, the U.S. has the highest and most active number of fact-checking organizations globally, relative to the African continent, which has either none or fewer than 3 fact-checking organizations in select countries[65]. For example, Meta has 11 certified fact-checking partners in both the U.S.[66] and India[67], despite India's user base of 400 million people being double the size of the U.S. user base. In further contrast, Meta has 6 fact-checking partners in Indonesia[68], serving close to 120 million monthly active users.

Fact-checking organizations need to be certified by the International Fact-Checking Network (IFCN) at Poynter— a U.S. media institute— to formally participate in the program at social media platforms. However,

the certification process can be a barrier for Global Majority partners because of lack of access, structure, resources and language, especially in authoritarian environments where formal documentation can lead to a high risk of persecution. Additionally, there are few quality controls to audit the veracity of debunked information by fact-checking partners, leading to a growing industry of "fake" fact-checkers. While the phenomenon was first brought to light during the Russia-Ukraine war[69], "fake" or pro-government fact-checking organizations have been a long-standing concern in Global Majority countries.

Automatic or human detection also may be of lower quality in Global Majority languages depending on human resources available and the quality of automated detection in those languages. Since organizations or external stakeholders need access to companies to be able to flag misleading content that may be eligible for enforcement, CSOs in the Global Majority may be at a disadvantage relative to U.S. CSOs with closer proximity to companies' leadership.

Evaluating platforms' effectiveness

Evaluating how well platforms combat misinformation requires understanding the extent to which companies remove, demote, or debunk content that makes false claims about elections and election-adjacent events. Some of the specific questions to help evaluate companies' preparedness are as follows:

1. Ongoing policies and processes:
   a. Is misinformation about elections removed, demoted, or flagged with warning labels in the country?
   b. How many fact-checkers and trusted partners (external organizations and stakeholders such as CSOs) are active in the country? Who are these fact-checkers and trusted partners and how reliable and unbiased are they?
   c. In addition to relying on fact-checking partners, what other mechanisms are in place for the social media platform to verify the veracity of election-related claims? Will there be an audit on the accuracy of fact-checking partners' debunking?
   d. How does each party—platform, civil society, and the regulator—define election misinformation, and what are the reasons for discrepancies in definitions?
   e. Has the platform organized training and discussions with civil society, trusted partners and fact-checking organizations ahead of the election? Has training covered how election observers, researchers, and civil society can escalate misinformation to dedicated teams?
   f. Does virality play a role in the extent to which misinformation is acted upon? Do companies act on election misinformation only when it goes viral across an entire country? Does the company monitor misinformation targeting specific communities in the country that may not have a large enough population to achieve virality nationally?

2. During the election period:
   a. Will there be dedicated staff that are responsible for escalating and responding to election misinformation in the country within each company? Are there ways

for local organizations to support platforms in the interpretation of context-dependent and localized content?
b. How can election observers, researchers, and civil society understand what is happening on the platform during an election period? Is data made available to external parties such as researchers?
c. Are there reports during election campaigns or assessments of past elections?
d. How do platforms respond if candidates make inaccurate claims about election results?

3. After the election period:
   a. How many false claims about elections were debunked or removed by the end of the election period? What types of claims were debunked or removed? How many escalations did trusted partners initiate and what were the outcomes?

Case Study: Election misinformation in Bangladesh

Bangladesh's most recent general election took place on January 7, 2024. Ahead of the prior general election of 2018, despite having 33 million monthly active users in the country, Facebook did not have any independent fact-checking partner in Bangladesh. There were no warning labels applied to content that made false claims about election results.

Over the month leading up to the 2018 election, "fake" news websites bearing logos of credible national news organizations began to emerge and were amplified through their social media channels[112]. In 2020, EU DisinfoLab published a report indicating that a sophisticated Indian disinformation campaign, backed by Srivastava Group, targeted international institutions to affect geopolitics in the region[113]. Of note is a Brussels-based think tank, South Asia Democratic Forum (SADF), that was reportedly tied to the disinformation campaign and organized an EU delegation to Bangladesh in the months leading up to the 2018 polls[114]. The same report indicated that SADF leveraged social media channels to amplify their support for the incumbent government at the time. Meta (formerly Facebook) and X (formerly

Twitter) removed several "fake" news accounts in the days leading up to the polls. However, civil society in the country alleges that the action was "too late", with misinformation and disinformation campaigns from these pages already having affected public opinion.

Public reports have indicated there are currently 7 established fact-checking organizations in Bangladesh, of whom three are Meta's partners. Fact-checkers have flagged spikes on political misinformation ahead of the 2024 polls,[115] but there is no public reporting available on whether platforms are proactively addressed misinformation about election integrity. Further, CSOs have documented sophisticated partisan networks sharing misleading, hyper-partisan content in the lead-up to the polls. Under existing policies, such content cannot be debunked by fact-checkers, and there are no supplementary interventions to reduce false hyper-partisan information that affects voter perceptions.

Appendix 3: Company rules on misinformation and election Integrity

YouTube

YouTube prohibits "certain types of misinformation that can cause real-world harm, like certain types of technically manipulated content, and content interfering with democratic processes."[70] This includes voter suppression, candidate eligibility, incitement to interfere with election processes, and distribution of hacked materials that may interfere with democratic policies. Specifically, false claims about certified election results in certain elections are prohibited (the 2021 German federal election; the 2014, 2018, and 2022 Brazilian Presidential elections)

Meta (Facebook, Instagram)

Meta removes content that could directly contribute to physical harm or interfere with democratic participation, as well as highly deceptive manipulative media.[71] In the context of an election, prohibited content includes misrepresentation of or misleading information about the dates, locations, times and methods for voting or voter registration and misrepresentation of who can vote, qualifications for voting, whether a vote will be counted and what information and/or materials must be provided in order to vote.[72] It is not explicitly stated whether these policies are uniformly applied across all general elections around the world, as well as regional or municipal elections. Meta uses artificial intelligence to detect new content that is likely to contain false information to send to third-party fact-checkers.[73]

TikTok

TikTok may remove content that violates its community guidelines, including "false claims that seek to erode trust in public institutions." This includes claims of voter fraud, misrepresentations of election dates, and attempts to intimidate voters or suppress voting.[74]

Bullying and Harassment

The problem

Bullying and harassment content refers to insults or slurs targeting individuals based on their personal attributes or protected characteristics (e.g., religious groups, racial groups). This differs from hate speech, which concerns attacks targeting an entire group of people based on protected characteristics. During elections, harassment often targets politicians, particularly women and those from marginalized groups, as well as citizens and activists engaging in political issues online. Aside from inherently being a harmful experience to those targeted, harassment can have the effect of suppressing political speech and civic participation, as it has historically driven women out of politics and deterred them from running for office[75].

Major social media platforms generally do not allow content containing prolonged insults, threats or dehumanizing language targeting individuals. Their bullying and harassment policies cover a wide range of targeted malicious content, especially those that aim to degrade, shame or portray people in a negative, often sexual light. These policies generally exist permanently, but in the context of elections, the problem may become especially pronounced in politicized contexts involving political figures or marginalized populations.

How do companies identify and address bullying and harassment?

Content that violates bullying and harassment policies is generally removed from major platforms. Since much of this enforcement is automatic, different languages will see a variation in the quality and precision of enforcement. Explicit forms of harassment that use common keywords will also be more effectively detected (slurs and common insults) than more subtle forms of bullying and harassment.

Policy and enforcement gaps

Common gaps across major internet companies are as follows:

- **Contextual meaning:** Policies often police content based on sexualization; however, what constitutes sexual varies by culture and region. In more conservative cultures, content depicting a fully clothed woman wearing bright red lipstick may be considered sexual, while in other cultures, more explicit sexualisation may be considered permissible.
- **Implicit attacks:** Implicit attacks and content combining images with text are more difficult for automatic systems to detect. This means that content that fits the definition of harassment may be undetected without specific attention to the implicit forms that this can take.
- **Enforcement across languages:** The effectiveness of enforcement of harassment policies depends on the quality of automated detection systems in a given language. Companies that have less data in languages spoken by smaller populations cannot train their detection systems to be as effective,

which then may result in gaps in enforcement. Because bullying and harassing language are highly contextual, unless detection systems are specifically trained to understand the specific context in a specific language, it is likely to result in enforcement gaps.

- **Harassment of politicians, public figures:** Companies consider insults targeting public figures to be part of free speech. This can result in gaps in enforcement for bullying and harassment targeted at public figures. Where safeguards are put in place for politicians and other civic leaders, it is important to understand if these are temporary or long-term.

Gaps in Global Majority countries

Countering bullying and harassment enforcement focuses on explicit content such as slurs that are relatively simple to detect automatically– as long as companies have the capability to detect them in a particular language. This means that implicit forms of harassment are unlikely to be detected unless proactive action is taken.

Evaluating platforms' effectiveness

Platforms' effectiveness in handling bullying and harassment during elections is dependent on their ability to incorporate regional and contextual considerations, and limitations on their policy enforcement. The following questions can aid in evaluation:

1. Regional and contextual considerations
   a. Is this policy applicable and enforced globally or limited to certain countries?
   b. How are regional or cultural contexts taken into account, including in automated systems?
   c. What is the process to identify and evaluate context and implicit language? How are CSOs included into this process?
   d. How is sexualization defined?
   e. How do the categories of "protected groups" change within the platforms' policies and systems during elections? What happens when new groups emerge?
   f. What resources are available for each of the countries to proactively moderate content, and are there sufficient resources available to escalate content?
   g. When there is internal disagreement on whether specific phrases or content should fall within the policy, what are the criteria and process for the company to come to a decision? Are local experts consulted?

2. Limitations on policy enforcement
   a. Are these policies enforced at all times or for a limited time around elections?
   b. How are "public figures" and "private individuals" defined by the platform, and what are the implications for enforcement?
   c. What types of harmful behaviors are responded to proactively versus reactively?
   d. Are there election-time enforcements that are more aggressive in a particular country?

Case Study: Bullying and Harassment in Pakistan

In Pakistan, women's electoral participation in national- and provincial- level elections remain low. In 2018, only eight women were elected to general seats while 61 were elected in reserved seat which is 20% of total representation[116]. Rights groups allege that women's participation is further negated with how they are bullied and harassed online[117]. According to a survey study by Digital Rights Foundation, 70 percent of Pakistani women are afraid of their photos being shared online, while 40% indicated they have been harassed or stalked on messaging apps[118]. In the lead up to the country's general election in 2019, multiple studies indicated that female politicians are more likely to receive objectifying, dehumanizing and sexualized comments across Internet platforms relative to their male counterparts[119]. Female activists and journalists reportedly experienced increasing amounts of online harassment and intimidation[120], including release of personal details, that have dire offline consequences[121]. In addition to disengaging from posting online, they can be forced out of jobs or barred from appearing in public forums, and in some extreme cases, can also be killed[122]

Meta's Bullying and Harassment policies do not allow severe attacks or sexualized commentary against public figures. Similarly, YouTube's policies do not allow prolonged insults or slurs based on an individual's intrinsic attributes, but provide a caveat to allow some degree of harassment against public figures[123]. These policies do not distinguish between power and societal dynamics within gender and sexual identities; therefore, they are insufficient in addressing the far-reaching consequences of attacks on women and gender minority communities relative to men. A 2016 case from Pakistan finds that a female public figure was murdered because of simply posting content online that her family believed would "dishonor" them[124]. While Meta prohibits sharing of non-consensual intimate images on its apps[125], what constitutes a "private" and "intimate" image has a range of context-specific definitions. A presumably harmless image of a woman and man at a park[126], or women at a protest, can trigger character assassination and harassment against them that could even lead to their murder, also known as "honor killing" in Pakistan. In November 2023, allegedly doctored photos of a young man and woman went viral on social media, resulting in them receiving death threats and being taken into police custody for safety[127].

Internet platforms do not provide any safeguards for non-consensual images of women posted online, nor apply context-specific definitions of sexualization. This poses significant risks particularly as female public figures became more active in the lead-up to the election in 2024.

Appendix 4: Company Rules on Bullying and Harassment

YouTube

YouTube's Harassment & Cyberbullying policy[76] does not allow content that contains prolonged insults or slurs based on someone's intrinsic attributes. These attributes include their protected group status, physical attributes, or their status as a survivor of sexual assault, non-consensual intimate imagery distribution, domestic abuse, child abuse and more. It offers additional protections to minors, specifically when it comes to content uploaded with the intent to shame, deceive or insult a minor. YouTube prohibits "Content that encourages abusive behavior, like brigading. Brigading is when an individual encourages the coordinated abuse of an identifiable individual on or off YouTube."

Meta (Facebook, Instagram)

Under its Bullying & Harassment policy[77], Meta does not allow threats and releasing personally identifiable information to sending threatening messages and making unwanted malicious contact. The policy distinguishes between private individuals versus public figures, as well as offers of additional safeguards for minors (under 18 years old) including removing claims about sexual activity, sexualisation of another adult, dehumanizing comparisons and bullying on the basis of physical characteristics. Meta also prohibits mass harassment and intimidation from multiple accounts, and removes coordinated efforts of mass harassment that target individuals at heightened risk of offline harm, such as victims of violent tragedies or government dissidents, even if the content on its own wouldn't violate their policies. Meta may remove objectionable content that is considered mass harassment towards any individual on personal surfaces, such as direct messages in inbox or comments on personal profiles or posts, but requires additional information or context to enforce this new policy.

TikTok

TikTok's Harassment & Bullying policy[78] prohibits "Promoting coordinated harassment of a person or attempting to create conflict between people, such as calling for others to flood comments with abusive language." TikTok does allow some content that is critical of public figures as "they are in a position of public attention and have ways to counter negative speech, and that the critique may be in the public interest to view." However, the company will remove content targeting public figures that violates other policies like threats, hate speech, sexual exploitation, doxxing and expressing a desire for someone to experience serious physical harm. TikTok's harassment and bullying policy does not explicitly reference brigading.

Coordinated Inauthentic or Deceptive Behavior

The problem

Coordinated inauthentic or deceptive behavior is a manipulative tactic of using fake, duplicated, and sometimes combined with authentic, social media accounts to operate an adversarial network on a social media platform, in order to harass, harm, or mislead

real users on the platform. Adversarial groups may do this for financial gain, such as spamming users using a network of fake accounts in order to sell goods; or for ideological purposes, such as by spreading disinformation campaigns through a number of accounts that appear to be genuine users. Both foreign and domestic information operations to interfere in elections may fall under companies' policies that prohibit these behaviors. Since adversarial networks can be responsible for spreading disinformation at a much larger scale than individual, unconnected individuals, disruption of whole networks can be a more effective way to target the root causes of disinformation, and also result in disrupting a much larger volume of disinformation content, compared to responding to individual pieces of content using the typical misinformation intervention methods described in the Misinformation section.

How do companies identify and act against coordinated inauthentic or deceptive behavior?

Coordinated inauthentic and deceptive behavior is generally prohibited by companies. To identify such behavior, they detect signals of coordination, such as indicators that an account is being operated from a location that doesn't match its stated location, or that multiple accounts are being operated from the same location, internet network, or device. Companies can also infer coordination by detecting behaviors on their platforms like duplicative content or accounts. Coordination could also involve a number of accounts encouraging harassing or other abusive behavior, also known as brigading, which is generally enforced under harassment policies.

Companies may detect these types of behaviors with automated systems, and also uncover deceptive efforts through manual investigations. When policy-violating types of inauthentic and deceptive behaviors are detected, the associated content and accounts are removed. Companies may publicize some of these takedowns in publicly-shared reports.

Policy and process gaps

- Platforms are not very transparent about their tactics and impact in this area, partly out of necessity of preventing bad actors from learning and adapting to their tactics to evade the companies' systems.
- Investigations into coordinated behavior require human effort, which can be costly and may mean that enforcement is incomplete and determined by prioritization.

Gaps in Global Majority countries

- Coordinated inauthentic or deceptive behavior policies were created to tackle foreign interference on elections or spam operations. Some types of malicious coordination by repressive governments or bad actors don't neatly fall under these or any other policies, leaving the platforms vulnerable to exploitation. (See case study below.)

- It is difficult to evaluate whether companies' efforts to enforce against coordinated harmful behavior, especially as related to elections, are

implemented during all elections in Global Majority countries.

- It is unclear if policies on coordinated deceptive and inauthentic behaviors are enforced automatically, or through manual investigations, which may result in limits to enforcement based on available resourcing.
- Companies are not transparent about how they `prioritize where and when they launch investigations.

Evaluating platforms' effectiveness

- How does each company become aware of coordinated efforts? Do the companies proactively detect or search for these behaviors? And if so, is done using with automated systems or with human investigators?
- How do these efforts vary by country? How do these efforts vary when there is an election?
- Do companies coordinate with government agencies, such as through the sharing of intelligence, to unearth and disrupt foreign influence operations on their platforms?
- Do platforms enforce on coordinated information operations, even by domestic operators and actors, and not just foreign actors?
- CSOs can draw attention to high risk accounts that are targeted by coordinated inauthentic or deceptive behavior.
- CSOs can seek transparency on whether platforms are making extra investments into preventing and disrupting coordinated inauthentic or deceptive behavior during elections in their country.

Case study: Coordinated malicious actors in Nicaragua

In 2020, the government of Nicaragua President Daniel Ortega used copyright law to have content by independent media outlets taken down from YouTube. After the government revoked the independent media outlet *100% Noticias'* broadcasting license in 2018, the outlet became reliant on its YouTube channels to hold its news archives and broadcast new reporting to the public, including on the government's repressive actions against citizen protestors. Only pro-government media outlets, which were largely controlled by the family members and allies of President Ortega, had access to interviews and events with state officials. Independent outlets like *100% Noticias* had to rely on using images and recordings of those sources in their own YouTube videos to report on events like presidential speeches. In March 2020, *100% Noticias* was

informed by YouTube that its two channels would be shut down due to reported violations of copyright. *100% Noticias* followed instructions in YouTube's notification email to file a "counter notice," stating the content was broadcast in the public interest, and constituted fair use, but received a notification that its claim could not be processed.[128] Another independent outlet, Confidencial, similarly had to challenge copyright violation allegations from YouTube. According to the Committee to Protect Journalists, the copyright complaints in both cases came from media outlets owned by the President's family members. This type of coordinated reporting does not typically fall under platforms' policies for coordinated deceptive behavior, since no inauthentic accounts were involved. These cases show how a platform can be vulnerable to malicious information operations exploiting system loopholes.

Appendix 5: Company rules on coordinated inauthentic or deceptive behavior

YouTube

Youtube's policy on spam and deceptive practices[79] prohibits "spam, scams, or other deceptive practices that take advantage of the YouTube community." While the publicly-posted policies do not specifically reference coordinated deceptive practices, Google's Threat Analysis Group[80] announces instances where YouTube channels are terminated as a result of their investigations into coordinated influence operations. In past elections, YouTube has publicized efforts to proactively monitor deceptive practices[81] including disinformation campaigns.

Meta (Facebook, Instagram, WhatsApp)

Meta defines Coordinated Inauthentic Behavior as when groups of people or Pages work together to mislead people about who they are, or what they're doing. They then takes down these networks for deceptive behavior, not for the content they're sharing. An example is a group that might misrepresent where in the world they are, either for financial gain or ideological purposes. Meta may take down these groups and accounts on both Facebook and Instagram through automated systems, or by manual investigation and removal. Meta publishes a quarterly Adversarial Threat Report[82] that provides information on major networks and threats that are taken down on Facebook and Instagram.

Additionally, Meta's Coordinating Harm[83]

policy prohibits the following behaviors that relate to elections:

- Calls for coordinated interference that would affect an individual's ability to participate in an official census or election
- Offers to buy or sell votes with cash or gifts
- Statements that advocate for, provide instructions on, or show explicit intent to illegally participating in a voting or census process.
- Content stating that census or voting participation may or will result in law enforcement consequences (for example, arrest, deportation or imprisonment).
- Statements of intent, support or advocacy to go to an election site, voting location, or vote counting location when the purpose of going to the site is to monitor or watch voters or election officials' activity using militaristic language (e.g. "war," "army," or "soldier") or an expressed goal to intimidate, exert control or display power (e.g. "Let's show them who's boss!," "If they're scared, they won't vote!").

WhatsApp[84] bans coordinated mass messaging and uses automated spam detection technology to spot "accounts engaging in abnormal behavior so they can't be used to spread spam or misinformation" and will ban accounts engaging in this behavior before users report them. Additionally, they "rely on machine learning to prevent accounts attempting to create groups at scale to message users," which is another coordinated abuse tactic.

TikTok

TikTok's Spam and Deceptive Account Behaviors Policy[85] prohibits "account behaviors that may spam or mislead its users. This includes conducting covert influence operations, and operating spam or impersonation accounts," with an exception for parody or fan-based accounts. Users are allowed to set up multiple accounts on TikTok "to create different channels for authentic creative expression, but not for deceptive purposes. [Tiktok does] not allow the use of multiple accounts to intentionally bypass [its] rules or their enforcement." Accounts that engage in these deceptive behaviors may be banned, as well as any new accounts created by the same users. TikTok disrupts covert influence operations with layered operations involving investigation, removal, and post-mortem analysis. The company reports the removal of any such networks in their Community Guidelines Enforcement Reports[86], during the quarter in which the full operations process has been completed.

Political Ads

The problem

Advertisements about political issues, election campaigns and candidates, and other political topics can be purchased by advertisers on a number of digital platforms, with various constraints. This type of paid content guarantees greater visibility among users than user-generated content. Advertisers can narrowly target the ads to users based on factors like age, gender, location, and context (e.g. topics). Users can choose options to see "fewer" or "less" of similar political ads as those they

see.Users can choose options to see "fewer" or "less" of similar political ads as those they see, but they cannot opt out of political ads altogether. Political ads can consist of either content created just for the advertisement, or a boost to "organic" or unpaid content, which can effectively give those who are able to pay a greater reach on these platforms. The same types of risky content and behavior among unpaid content may also occur with paid content; though one key difference with unpaid content is that companies generally create checkpoints to review advertising content before it is published, whether through automated or human review processes.

How do companies address issues in political ads?

Some internet companies do not allow political ads on their platforms at all, such as TikTok. While this prohibition prevents any potential negative effects from political ads, organizations that promote voter participation also cannot use political ads as a tactic for growing voter engagement.

Companies such as Google and Meta that do allow digital advertising related to elections, political campaigns, and social issues are subject to laws and regulations for these types of ads in many countries. Other than legal requirements, companies also have their own policies on what types of elections and politics-related ads are and are not allowed. Certain types of harmful content are generally banned from ads by the companies, such as hate speech, promotion of hate groups, or content that deters or suppresses voter participation, but companies do not necessarily ban ads containing misinformation. These companies may also place requirements on advertisers to register or verify their identity in order to be able to run political or election ads. This enables the platforms to verify that advertisers are located in the countries in which they are running ads, and to display disclosures to users on who paid for an ad. Similar to unpaid content, companies utilize a combination of automated systems and humans to detect and reject ads content that violates their policies. Major platforms provide some transparency by making ads that are run on their platforms available to the public through a searchable database called an ads library or ads transparency center. As of the fall of 2023, both Google and Meta created new policies requiring advertisers to disclose when ad content is manipulated or generated to inauthentically depict real or realistic-looking people or events, such as with the use of artificial intelligence.

Gaps in Global Majority countries

Meta requires advertisers in over 220 countries and territories to authorize and disclose who they are, in order to run electoral, or political ads on Facebook and Instagram.[87] In around 40 additional countries, mostly outside the Global Majority, the same requirement applies if advertisers run ads on social issues. For countries with such authorization and disclosure requirements, Meta proactively detects or reactively reviews ads for compliance with requirements. Google requires advertiser verification in only 11 countries, and requires compliance with additional restrictions on political ads in an additional 7 countries.[88] This means that in most

countries in the Global Majority, users likely cannot know who is paying for political and election ads. YouTube ads are subject to the same Community Guidelines as non-paid content, which means voter suppression elements are not allowed in either paid or non-paid content. However, it is unclear if YouTube prioritizes enforcement on paid voter suppression content, which likely has greater reach and visibility than non-paid content. Additionally, YouTube does not require disclosing the use of synthetic content in regions where advertisers are not required to provide verification in order to run election ads, which applies to most countries in the Global Majority.

Generally, "political" and "social issues" could be broad categories. Platforms are not transparent about how they define these categories, and how accurately they can identify or detect these types of content in different languages and cultural contexts. This can be inferred based on a retroactive review of ads that have already run and are available in ads transparency centers. Publicly available materials do not make it clear in what types of elections and in which countries platforms will proactively enforce political and election ads for compliance with all policies, versus reactively review them.

Evaluating internet platforms' performance & preparedness for elections

1. What are the limits on political advertising in your country that apply to digital ads (e.g. blackout periods and limitations by topic or by entity purchasing the ads)?

2. If political ads are allowed in your country, how is "political" defined?

3. Are ads that question or undermine election processes, institutions, or results prohibited in your country? Do platforms allow ads that incorrectly state or cast doubt on the results of past elections that are widely accepted?

4. What checks do platforms have in place to limit or prohibit ads that involve rumors or misleading information about political candidates or political figures that can influence the outcome of an election or cause potential harassment or harm to those figures?

5. Are ads that are run in your country available to researchers and the public to view in an ads library or similar interface that provides transparency? Is there a data source available for you to understand what's happening with ads on the platform in real time?

Case Study: Political ads in Bangladesh

Under the previous Awami League-led government, opposition parties and their members in Bangladesh faced challenges when running political ads on social media due to strict rules set by platforms. While these rules are designed to ensure transparency and prevent foreign interference in elections, they particularly created complexities for opposition parties, in many cases due to asset freeze orders issued by the previous government. Additionally, social media platforms like Meta enforce a verification process by checking the authenticity and the location of the payer. This verification process put opposition parties at a disadvantage, especially since many of their members and leaders, including the acting vice chairperson of the Bangladesh Nationalist Party, live abroad due to political security reasons, and campaign funds often originate from outside the country. Further, providing the specific address that can be verified using government documents becomes difficult because of safety reasons. As a result, meeting the stringent verification requirements has been historically difficult for opposition parties trying to make ad payments from overseas.

An unequal distribution of political advertising spending could also be seen, as the ruling Awami League ran political ads on the social media platforms and benefited from greater exposure and visibility on social media, which further amplified the disadvantages faced by opposition entities in reaching and engaging with electorates. Ultimately, the dynamics of political communication were tilted in favor of the now fallen Awami League.

Appendix 6: Company rules on political ads

Meta (Facebook, Instagram, WhatsApp)

According to Meta's advertising policies[89], advertisers can run ads on Facebook and Instagram about social issues, elections, or politics, provided the advertiser complies with all applicable laws and the authorization process required by Meta. Meta may restrict issue-based, electoral or political ads. In addition, certain content related to elections may be prohibited by local law or removed in specific regions ahead of voting. For instance, ads that undermine elections and voting in specific ways are prohibited in the U.S., Italy, Brazil, and Israel[90]. Meta also prohibits ads that include content debunked by third-party fact checkers. Meta provides all ads that have been run on its platforms in an Ads Library[91], along with related information, stored for up to 7 years. As of this year, advertisers now have to disclose when their political or social issue ads are "digitally created or altered" through the use of Artificial Intelligence and contain "a photorealistic image or video, or realistic sounding audio."[92] Information about digitally altered ads will be captured in the Ad Library. As of

November 2022, advertisers running political, elections, and social issues ads are barred from using Meta's generative AI advertising tools.[93]

WhatsApp currently does not allow political candidates and political campaigns to use the WhatsApp Business Platform[94].

YouTube

Ads running on YouTube are subject to Google Ads policies, content that lives on the platform are subject to YouTube Community Guidelines, and channels that are part of the YouTube Partner Program are subject to YouTube Monetization policies.[95] Google's political and election advertising requirements vary by region, and advertisers are expected to comply with any local legal requirements such as election advertising and campaigning "silence periods" for any geographic areas they target.[96] Some regions also have specific restrictions and prohibitions such as limiting how election ads can be targeted. In around 12 regions/countries, advertisers are required to verify in order to run election ads.[97] In regions where election advertiser verification is required, advertisers must prominently disclose when their ads contain synthetic content that "inauthentically depicts real or realistic-looking people or events," with disclosures displayed clearly and conspicuously, on relevant images, video, and audio content. Advertisers are not allowed to misrepresent or conceal their country of origin to create ads about politics, social issues, or "matters of public concern" that are directed at users in a country other than the advertiser's country. Ads run on Google platforms are available in the Ads Transparency Center[98], which contains an interface for viewing political ads[99] run in the countries where advertiser verification is required.

TikTok

TikTok's Advertising Guideline state that political or issue-based advertising are not allowed[100]. They define these categories as[101]:

1. Candidates or nominees for public office, political parties, and elected or appointed government officials are prohibited from advertising.

2. The spouses of candidates, elected, or appointed government officials with official duties or offices are prohibited.

3. Royal Family members with official government capacities are also prohibited.

However, government entities may be able to advertise if working with a TikTok Sales Representative.

How This Guide Was Created

This overview is based on research on publicly available information on companies' policies, product launches during elections, and enforcement activities and impact. This information includes community guidelines, newsroom posts, threat analysis reports, and transparency reports published by the companies. Additionally, we reviewed reports by think tanks, civil society groups and human rights organizations, as well as press stories from credible news organizations, for country-specific case studies and examples. We complement secondary research with interviews and consultations with civil society experts in multiple Global Majority countries.

Appendix 7

Detailed View: Election Protections by Platform

Tool	Availability per platform		
	Meta	Youtube	Tiktok
Election Integrity and Voter Participation			
Voting information hub or center	US only		US only
Voting alert product (day-of reminders, voter registration)	Yes		
In-feed voting notification in multiple languages	Yes		
Election results product	Limited countries		US only
Contextual friction product (are you sure you want to share this)	Yes	Yes	Yes
Voter suppression policy	Yes	Yes	Yes
Intimidation policy	Yes	Yes	Yes
Incitement to violence policy	Yes	Yes	Yes
Countering Misinformation			
Fact checking program	Yes	Yes	Yes
Trusted partner / flagger program	Yes	Yes	Yes
Media/digital literacy program	Yes	Yes	Yes
Manipulated media disclosure policy	Yes	Yes	Yes
Preventing Interference			
Coordinated inauthentic/ deceptive behavior policy	Yes	Yes	Yes
Covert influence network disruptions	Yes	Yes	Yes
Transparency and Accountability			
Political advertising allowed	Yes	Yes	No
Handling government requests	Yes	Yes	Yes
Disclosure on state-controlled media accounts & content	Yes	Yes	Yes
Treatment of proscribed entities (Taliban, Hamas, Tatmadow)			

Tool	Availability per platform		
	Meta	Youtube	Tiktok
Transparency and Accountability			
Data access for researchers	Yes	Yes	Yes
Data access for journalists			
Safeguarding Candidates and Key Civic Figures (Election commissioners, election observers, activists, journalists)			
Verified badge	Yes	Yes	No
Harassment policy	Yes	Yes	Yes
Brigading policy	Yes	Yes	
Impersonation policy	Yes	Yes	Yes
Account protection & security program for high profile accounts	Yes		

Endnotes

1 The Economist (2023) *2024 is the biggest election year in history*, The Economist. (Accessed: 04 December 2023).
2 Social Internet is an umbrella term used to define social media and messaging platforms (apps, websites and/or services) that facilitate user-to-user communication on the Internet.
3 Dumbrava, C. (2021) *Key risks posed by social media to democracy - European Parliament*, *https://www.europarl.europa.eu/*. (Accessed: 04 December 2023).
4 Fujiwara, T., Müller, K., & Schwarz, C. (2023). The effect of social media on elections: Evidence from the United States. *Journal of the European Economic Association*, jvad058. See also an overview of a series of academic studies in collaboration with Meta: https://www.nyu.edu/about/news-publications/news/2023/july/2020-election-studies-reveals-power-of-facebook--instagram-algor.html.
5 Meta Platforms Inc. (2023) *Second Quarter 2023 Results*. *https://investor.fb.com* (Accessed: 04 December 2023).
6 Tworek, H. (2021). Facebook's America-centrism is now plain for all to see. *Centre for International Governance Innovation*.
7 *ibid*
8 https://democracyfund.org/idea/the-growing-movement-for-platform-accountability/
9 Sarah T. Roberts, Behind the Screen: Content Moderation in the Shadows of Social Media (Yale University Press, 2019), 58, 169; Sarah Myers West 'Raging Against the Machine: Network Gatekeeping and Collective Action on Social Media Platforms' (2017) 5(3) Media and Communication 28, 28.
10 "Adding a 'D' to the ABC disinformation framework," Alaphilippe, A (2020). https://www.brookings.edu/articles/adding-a-d-to-the-abc-disinformation-framework/
11 https://www.cnbc.com/2022/11/07/social-media-platforms-prep-for-election-day-misinformation.html
12 https://www.nytimes.com/2022/08/23/technology/midterms-misinformation-tiktok-facebook.html
13 https://electionlab.mit.edu/articles/building-voter-trust-social-media
14 https://about.fb.com/news/2022/08/meta-plans-for-2022-us-midterms/
15 https://blog.youtube/news-and-events/the-2022-us-midterm-elections-on-youtube/
16 https://newsroom.tiktok.com/en-us/our-commitment-to-election-integrity
17 https://datareportal.com/reports/digital-2023-philippines
18 https://www.theguardian.com/technology/2019/feb/01/world-internet-usage-index-philippines-10-hours-a-day
19 https://www.bloomberg.com/news/features/2017-12-07/how-rodrigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook
20 https://www.hrw.org/news/2023/01/26/philippines-officials-red-tagging-indigenous-leaders-activists
21 https://internews.org/wp-content/uploads/2023/01/Red-Tagging-in-the-Philippines.pdf
22 https://www.nationalgeographic.com/magazine/article/in-this-country-social-media-is-successfully-rewriting-an-autocratic-past
23 https://newsroom.tiktok.com/fil-ph/helping-filipinos-make-informed-decisions-during-the-2022-philippine-elections
24 https://about.fb.com/news/2022/04/philippines-2022-general-election
25 https://newsroom.tiktok.com/fil-ph/helping-filipinos-make-informed-decisions-during-the-2022-philippine-elections
26 https://mb.com.ph/2022/03/03/google-youtube-support-philippine-elections
27 https://www.wired.com/story/youtube-philippines-election/
28 https://creativegood.com/blog/22/concentration-of-power.html
29 https://www.sciencedirect.com/science/article/abs/pii/S0167624501000671
30 https://www.atlanticcouncil.org/in-depth-research-reports/report/scaling-trust
31 Roberts (n 13) 93-94; Gillespie et al (n 3) 20-21; Witt, Suzor and Huggins (n 8) 575; Edoardo Celeste et al, The Content Governance Dilemma Digital Constitutionalism, Social Media and the Search for a global Standard (Palgrave Macmillan, 2023) 15-16.
32 Tarleton Gillespie et al, 'Expanding the debate about content moderation: scholarly research agendas for the coming policy debates' (2020) 9(4) Internet Policy Review 1, 3.
33 https://bhr.stern.nyu.edu/tech-content-moderation-june-2020
34 https://paperswithcode.com/task/cross-lingual-transfer
35 https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis

36 https://www.ndi.org/elections-calendar-all?year=2023
37 https://newsroom.tiktok.com/en-africa/staying-committed-to-election-integrity-ahead-of-the-nigerian-general-election
38 https://about.fb.com/news/category/election-integrity
39 https://blog.twitter.com/en_us/tags.blog--elections
40 https://about.fb.com/news/2023/11/how-meta-is-planning-for-elections-in-2024
41 https://scontent-iad3-1.xx.fbcdn.net/v/t39.8562-6/409833832_904454820678505_4813088771997457589_n.pdf?_nc_cat=102&ccb=1-7&_nc_sid=e280be&_nc_ohc=lIyUVXJgeqwAX80gMly&_nc_ht=scontent-iad3-1.xx&oh=00_AfBmjTKveicahPQ1fmnVtAKt-LsuShUJJBSjqqRM-m5QFg&oe=6580F6BD
42 https://blog.youtube/inside-youtube/us-election-misinformation-update-2023
43 https://blog.google/around-the-globe/google-europe/supporting-elections-for-european-parliament-2024/
44 https://blog.google/technology/ai/our-responsible-approach-to-building-guardrails-for-generative-ai/
45 https://about.fb.com/news/2023/11/how-meta-is-planning-for-elections-in-2024
46 https://about.fb.com/news/2024/02/mcas-whatsapp-helpline-curbing-the-spread-of-ai-generated-misinformation/
47 https://www.facebook.com/government-nonprofits/2024-south-african-general-elections
48 https://blog.google/outreach-initiatives/civics/how-were-approaching-the-2024-us-elections
49 https://blog.google/around-the-globe/google-europe/supporting-elections-for-european-parliament-2024
50 https://blog.google/intl/en-in/company-news/outreach-initiatives/supporting-the-2024-indian-general-election
51 https://blog.google/technology/ai/our-responsible-approach-to-building-guardrails-for-generative-ai/
52 https://blog.google/intl/en-in/products/platforms/our-approach-to-responsible-ai-innovation-on-youtube/
53 https://newsroom.tiktok.com/in-id/komitmen-tiktok-melindungi-integritas-pemilihan-umum-jelang-pemilu-di-indonesia
54 https://newsroom.tiktok.com/en-africa/zaelections
55 https://newsroom.tiktok.com/es-latam/elecciones-mexico-2024
56 https://newsroom.tiktok.com/en-eu/our-work-to-prepare-for-the-2024-european-elections
57 https://newsroom.tiktok.com/en-us/protecting-election-integrity-in-2024
58 https://newsroom.tiktok.com/en-us/our-commitment-to-election-integrity
59 https://about.fb.com/news/2022/08/meta-plans-for-2022-us-midterms/
60 https://blog.youtube/news-and-events/the-2022-us-midterm-elections-on-youtube/
61 https://blog.youtube/news-and-events/the-2022-us-midterm-elections-on-youtube/
62 https://apnews.com/article/election-workers-threats-trump-georgia-indictment-5b056e2c97b-fd7146b3bd19cf7f9f588
63 Meta Platforms Inc (2023) *Transparency center: Approach to newsworthy content, https://transparency.fb.com*.
64 www.france24.com/en/live-news/20221116-india-fact-checkers-face-threats-jail-in-misinformation-fight
65 https://reporterslab.org/fact-checking/
66 https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking/partner-map
67 https://about.fb.com/news/2022/07/expanding-our-third-party-fact-checking-program-in-india/
68 https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking/partner-map
69 Estrin, D., O'Connor, G. and Fox, K. (2022) *Who's checking the fact checkers?*, *NPR*. Available at: https://www.npr.org/ (Accessed: 05 December 2023).
70 https://support.google.com/youtube/answer/10835034?hl=en&ref_topic=10833358
71 https://transparency.fb.com/en-gb/policies/community-standards/misinformation/
72 https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/
73 https://ai.meta.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/
74 https://www.tiktok.com/safety/en/election-integrity/
75 https://www.csis.org/analysis/against-odds-overcoming-online-harassment-women-politics
76 https://support.google.com/youtube/answer/2802268?hl=en#:~:text=We%20don't%20allow%20content,behaviors%2C%20like%20threats%20or%20doxxing.
77 https://transparency.fb.com/en-gb/policies/community-standards/bullying-harassment/

78 https://www.tiktok.com/community-guidelines/en/safety-civility/
79 https://support.google.com/youtube/answer/2801973?hl=en&ref_topic=9282365
80 https://blog.google/threat-analysis-group
81 https://blog.google/outreach-initiatives/civics/our-ongoing-work-to-support-the-2022-us-midterm-elections/
82 https://about.fb.com/news/tag/coordinated-inauthentic-behavior/
83 https://transparency.fb.com/policies/community-standards/coordinating-harm-promoting-crime/
84 https://transparency.fb.com/policies/community-standards/coordinating-harm-promoting-crime/
85 https://www.tiktok.com/community-guidelines/en/integrity-authenticity/#6
86 https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/
87 https://www.facebook.com/business/help/2150157295276323?id=288762101909005
88 https://support.google.com/adspolicy/answer/6014595?hl=en#zippy=%2Cunited-states-us-election-ads%2Cphilippines
89 https://www.facebook.com/business/help/253606115684173
90 https://www.facebook.com/business/help/253606115684173
91 https://transparency.fb.com/researchtools/ad-library-tools
92 https://about.fb.com/news/2023/11/how-meta-is-planning-for-elections-in-2024/
93 https://www.facebook.com/business/help/297506218282224?id=649869995454285
94 https://faq.whatsapp.com/518562649771533
95 https://www.youtube.com/howyoutubeworks/our-commitments/supporting-political-integrity/#political-advertising
96 https://support.google.com/adspolicy/answer/6014595?sjid=580792505112816630-NA#zippy=
97 https://support.google.com/adspolicy/troubleshooter/9973345?hl=en
98 https://adstransparency.google.com
99 https://blog.google/outreach-initiatives/civics/our-ongoing-work-to-support-the-2022-us-midterm-elections/
100 https://www.tiktok.com/creators/creator-portal/en-us/community-guidelines-and-safety/tiktoks-stance-on-political-ads/
101 https://ads.tiktok.com/help/article/tiktok-advertising-policies-industry-entry?redirected=2
102 https://www.premiumtimesng.com/news/586788-nigeriadecides2023-only-27-of-eligible-voters-decide-who-becomes-nigerias-president.html
103 https://ng.boell.org/en/2022/10/26/bots-and-biases-role-social-media-nigerias-elections
104 Bello, A. W., & Kaufhold, K. (2023). Power to the People: Social Media as a Catalyst for Political Participation in Nigeria. *International Journal of Interactive Communication Systems and Technologies (IJICST)*, *12*(1), 1-17.
105 https://www.bbc.com/news/world-africa-66647768
106 https://www.bbc.co.uk/news/world-africa-66727354
107 https://www.nytimes.com/2023/09/06/world/africa/nigeria-decision-presidential-election.html
108 https://about.fb.com/news/2023/02/how-meta-is-preparing-for-nigerias-2023-general-elections/
109 https://about.fb.com/news/2023/02/how-meta-is-preparing-for-nigerias-2023-general-elections/
110 https://www.gatescambridge.org/our-scholars/blog/how-cambridge-analytica-influenced-nigerias-elections
111 https://newsroom.tiktok.com/en-africa/staying-committed-to-election-integrity-ahead-of-the-nigerian-general-election
112 Dhaka Tribune (2018) *Fake news hits Bangladeshi news sites before polls.* www.dhakatribune.com/bangladesh/election/161200/fake-news-hits-bangladeshi-news-sites-before-polls (Accessed: 05 December 2023).
113 EU DisinfoLab (2020) *Indian chronicles: Deep Dive into a 15-year operation targeting the EU and UN to serve Indian interests*. Available at: https://www.disinfo.eu (Accessed: 05 December 2023).
114 Bergman, D. (2022) *How a pro-india 'disinformation' group helped the Awami League government*, Netra News. (Accessed: 05 December 2023).
115 Islam, Z. and Khan, M.J. (2023) *Spin doctors go into overdrive ahead of polls*, The Daily Star. Available at: https://www.thedailystar.net (Accessed: 05 December 2023).
116 Violence Against Women in Elections in Pakistan, Centre for Peace and Development Initiatives, ISBN:978-969-2227-24-7, pg-15, https://pakvoter.org/wie/wp-content/uploads/2022/07/VAWE.pdf , accessed 14th December, 2023.

117 https://cpj.org/2020/09/as-ruling-party-fans-spew-online-abuse-pakistans-female-journalists-call-for-government-action/
118 https://digitalrightsfoundation.pk/wp-content/uploads/2017/12/UNSR-Submission-by-DRF.pdf
119 https://digitalrightsfoundation.pk/wp-content/uploads/2019/01/Booklet-Elections-Web-low.pdf
120 https://cpj.org/2020/09/as-ruling-party-fans-spew-online-abuse-pakistans-female-journalists-call-for-government-action/
121 https://www.hrw.org/news/2020/10/22/online-harassment-women-pakistan
122 https://thediplomat.com/2022/07/honor-killings-continue-unabated-in-pakistan/
123 https://support.google.com/youtube/answer/2802268?hl=en
124 https://www.dawn.com/news/1271213
125 https://about.meta.com/actions/safety/topics/bullying-harassment/ncii
126 https://www.cbsnews.com/news/honor-killing-pakistan-doctored-photo-woman-boyfriend-goes-viral-arrests/
127 https://www.bbc.com/news/world-asia-67551554
128 https://cpj.org/2020/05/youtube-censor-nicaragua-outlets-100-noticias-confidencial-ortega/

TECHGLOBAL INSTITUTE

techglobalinstitute.com