



June 12, 2023

National Telecommunications and Information Administration
United States Department of Commerce
Request for comment: *AI Accountability Policy Request*

To whom it may concern:

Thank you for the opportunity to provide recommendations for accountability measures and policies related to Artificial Intelligence.

While experts and academics have identified risks of generative AI to democracy and political processes, an application of a gendered lens is notably missing. Gendered disinformation is spread utilizing misogynistic tropes and gendered stereotypes, and can lead to offline targeting and violence. Foreign actors have increasingly employed gendered disinformation as a tactic to exert influence in other countries. Thus, the use of generative AI technologies to bolster these campaigns represents a serious threat to national security as well as representative democracy.

At #ShePersisted, we have spent several years [documenting](#) the way women in politics and in journalism experience relentless [volumes of online abuse](#), threats, and [gendered disinformation](#) campaigns on social media platforms. These campaigns are designed to discredit, devalue, and delegitimize women's political standing, with the goal of ultimately undermining their ability to participate in civic life.

Historically, technologies have been designed largely without explicit consideration of gender, with regulators failing to recognize and address its impact. Gender-based violence online also intersects with issues of racial bias, targeting of religious minorities and gender identity. The lack of oversight to the way such communities are targeted online has resulted in the growth of gender bias which underlines the equal participation and free expression of women. The rise of generative AI technology simply exacerbates issues of gender bias online. It is therefore essential to recognize that action is needed to hold digital platforms accountable for addressing current online harms. Social media companies must be held responsible for facilitating platform harms, whether AI-generated or not.

It is within the context of understanding the ecosystem of how gendered disinformation is organized and amplified that informs the following recommendations.

Gender-informed recommendations for AI policies:

1. Intensify the resolve to meaningfully address gendered disinformation to ensure election integrity in advance of 2024 elections across the globe.

Approaches to regulating AI technologies must first be grounded in an understanding of how technology-facilitated gender-based online violence and gendered disinformation is currently manifesting. Gender bias in politics is sadly nothing new – yet the way digital platforms are designed exponentially increases misogyny to the point of weakening social norms of inclusion and civil discourse, and normalizing abuse and impunity for its perpetrators. Social media algorithms alter behavior by incentivizing fake and sensationalized content – over-supplying it to users in the name of profit. This makes it much easier for gendered disinformation campaigns against women leaders to be organized, amplified, and cheaply financed, reaching millions of people.

Attacks on women and the use of gendered disinformation is particularly acute during periods of political transition and elections. In 2024, there will be more than 60 federal elections across the globe with more than 2 billion people eligible to vote for the first time in very fragile environments where illiberal actors are able to [utilize the tools](#) provided by [social media companies](#), largely with impunity. Faster AI tools for [election-related communication and messaging](#) could have a profound impact on how voters, politicians, and reporters see candidates, campaigns and those administering elections.

The federal government should take increased responsibility for studying gendered disinformation and its implications for women vying for public office and holding positions such as Secretary of State and heads of Election Management Bodies before the 2024 elections, as gendered disinformation is also aimed at undermining elections and nonpartisan administrators and volunteers.

Awareness of gendered disinformation has grown in recent years as the empirical evidence of its impact mounts, yet it is not acted on meaningfully in the regulatory sphere in the United States. Academics, independent researchers, and nonprofit civil society organizations have mapped how [influence operations](#) on gendered disinformation are coordinated. **This baseline understanding about the impact of technology on women in public life and gender more broadly should be the foundation of exploring emerging AI technologies.**

AI applications that allow users to generate new text, audio, visual, and audio-visual content will strengthen gendered disinformation, particularly if the ability to identify AI-generated content remains weak. This poses a developing threat to women, as such content can be amplified on social media in ways similar to existing attacks - “amping up at unprecedented speed, scope and scale of that chronic abuse that women are already facing,” [according to Gina Neff](#), Executive Director of the Minderoo Centre for Technology & Democracy at the University of Cambridge.

It is important to recognize that the lack of federal regulation in the United States creates difficulty for victims of online abuse to hold platforms financially responsible. Regulating or placing certain kinds of liability on digital platforms is appropriate and necessary, given their

role in facilitating technology-facilitated gender based violence. Other jurisdictions such as Canada, Australia, Germany, France, and the European Union are exploring such measures.

While limitations for curtailing harms due to the advantages given to platforms by Section 230 is unique to the United States, this should not stall serious investment into other forms of accountability mechanisms and the exploration of measures to prevent or mitigate online harms based on gender.

For example, many countries have enacted laws specifically targeting [image-based abuse](#) to address the non-consensual sharing or distribution of intimate images or videos without the subject's consent, often with the intention of causing harm, embarrassment, or harassment. While specific provisions and penalties vary from one jurisdiction to another, exploring legal frameworks surrounding image-based abuse should be intensified immediately in the context of the growth of AI, outlined by Cynthia Khoo in [Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence](#), such as: (1) criminal laws to make it illegal to share or distribute intimate images without consent, (2) civil remedies to victims of image-based abuse with a means to protect their privacy, seek damages, and restore their reputation, (3) legislation and legal reforms designed to close gaps in existing legislation or update legal frameworks to better reflect the digital age, (4) international efforts - some countries have ratified or endorsed international conventions or treaties that address online harassment, cybercrime, or violence against women, which can indirectly encompass image-based abuse.

There is no shortage of policy or regulatory ideas for addressing gender-related online harms. What's missing is a political will to act on existing guidelines and implement policies that meaningfully hold social media companies to account for facilitating the spread of harmful gendered content on their platforms.

AI developers should also be held to account for facilitating the creation of harmful gendered content. Policymakers and regulators must be emboldened by an understanding that technology-facilitated gender-based violence and gendered disinformation impacts not only individuals and society, but the foundations of democracy. Action taken to improve on existing regulatory shortcomings must be immediate, to both address current harms and to prevent the weaponization of emerging AI technologies by malign actors.

2. Form a task force designed to press for greater transparency and regulatory solutions among mainstream social media platforms operated by Meta, Alphabet, Twitter, and ByteDance.

At minimum, the United States should require transparency from platform companies regarding their content moderation policies and decisions, as well as the outcomes of such policies and decisions concerning technology-facilitated gender-based violence and gendered disinformation. Without more and better data providing insights into how such policies and decisions are made and implemented, governments, regulators, and the public will be stymied in influencing how the general public is governed and affected by digital platforms and the use of AI.

With additional resources, a gendered disinformation and AI task force could be embedded within the [Global Partnership for Action on Gender-Based Online Harassment and Abuse](#), a global body comprised of 12 governments that brings together international organizations, civil society, and the private sector to better prioritize, understand, prevent, and address the growing scourge of technology-facilitated gender-based violence.

3. Invest in research and development of AI tools that can detect and combat gendered disinformation.

Existing digital tools already pose a considerable threat to women. Through algorithms designed to foster gender bias, social media platforms that prioritize profit over the protection of users are benefiting from unique online harms such as image-based abuse and doxxing.

Sensity AI estimates that between 90% and 95% of all deepfakes generated concern image-based abuse. The vast majority of those deepfakes, about 90%, are targeted at women. AI technologies are rapidly improving and can now generate realistic visuals solely with text input. Per the [White House's Blueprint for an AI Bill of Rights](#), media generated with “AI-enabled ‘nudification’ technology” is becoming “increasingly realistic” and “difficult for both humans and AI to detect as inauthentic.” The content generated by these applications can also be non-sexual but still harmful; for example, Vice President Kamala Harris and Representative Nancy Pelosi have both been featured in manipulated videos falsely depicting them as drunk or making fabricated controversial statements.

Image generators and voice generators can be used with similar intentions. Text-based applications can quickly generate false texts in the style of a social media post or news article, and can be trained to regurgitate false information about an individual. These emerging tools provide a new opportunity for weaponization by both domestic and foreign actors seeking to influence voters, target women leaders, and achieve their own malign aims.

Additional resources should be allocated to initiatives such as the [National Artificial Intelligence Research Resource](#) with a direction to focus on gendered disinformation specifically, and how it is perpetrated utilizing emerging AI technologies. [Existing investments made by Congress](#) into AI R&D can also be further expanded to include research into the extent of gendered disinformation and ethical considerations pertaining to gender-based harms. These resources should be directed towards developing tools to identify gendered disinformation in text, audio, audio-visual and image-based content generated by emerging AI technologies, and towards research encompassing listening and trend monitoring, the mapping and tracking of malign actors, and analysis of terms of service compliance. Doing so will ensure audits, policies, and risk and impact assessments can be properly tailored to mitigating gendered disinformation predicated upon a foundational understanding of the threat, setting the stage for continued long-term analysis.

While there is an overreliance on media literacy as a remedy for curtailing mis/disinformation, AI might also be helpful in assessing to which extent digital and media literacy games for educational purposes aimed at consumers may (or may not) be able to address disinformation, therefore refining our set of existing response tools.

4. Consider gender balance and representation of expert communities who are targeted by gender-based violence online when setting up new advisory boards, developing audits and assessments.

Governments must work to ensure audits and regulatory efforts are underpinned by an understanding of how the misuse of any technology manifests uniquely when targeting women, and how emerging generative AI technologies can be used to perpetrate new attacks based on preexisting trendlines. To ensure women's perspectives are included in discussions of AI bias and image-based abuse, as well as in the development of regulatory frameworks to address online harms, we recommend:

- *Representation in decision-making bodies:* Actively involve women experts, researchers, and advocates in decision-making bodies, task forces, and committees focused on AI bias, image-based abuse, and online harm regulation. This ensures that diverse viewpoints are considered and helps prevent biases and blind spots.
- *Collaboration with women-led organizations:* Engage and collaborate with women-led organizations, advocacy groups, and nonprofits working on issues related to AI bias, image-based abuse, and online harm, particularly those which are not receiving funds from Big Tech and are therefore entirely free from their pressure. These organizations often have valuable insights and expertise to contribute to the discussions and can help identify and address specific concerns faced by women.
- *Research and data collection:* Encourage research and data collection that specifically examines the gendered impacts of AI bias, image-based abuse, and online harm. By understanding the unique challenges faced by women in these areas, appropriate solutions and regulations can be developed.
- *Empowerment through education and awareness:* Provide educational resources and awareness campaigns that specifically showcase the tactics of the perpetrators of gendered disinformation campaigns, highlighting their manipulation of AI, image-based abuse, and online harm. Empower citizens to recognize and report incidents, and encourage their active participation in shaping solutions and regulatory frameworks.
- *Inclusive policy development:* Ensure that policy development processes related to AI bias, image-based abuse, and online harm are inclusive and participatory. Seek input from women experts, organizations, and affected communities during the formulation and review of policies. This will help address the specific challenges faced by women and avoid the perpetuation of biases or blind spots.
- *Engaging tech companies:* Encourage tech companies to prioritize gender diversity and inclusivity in their development of AI systems and online platforms. Advocate for the implementation of responsible AI practices, including robust testing for biases, transparency in algorithms, and mechanisms to address image-based abuse and online harassment.

By actively involving women in these discussions, regulatory frameworks can be developed with a better understanding of the gendered implications and a focus on protecting women's rights and safety online.

In sum, considerations of gender must be kept at the forefront when designing regulatory measures to prevent the further exacerbation of societal divisions and to ensure women can

freely express themselves without fear of retaliation. This is an important moment of opportunity as the landscape of AI widens and a global race to regulate generative AI ensues. If threats posed to women by emerging AI technologies go unmitigated, progress made towards gender equality will be reversed, democracy undermined, and women's ability to participate in the public sphere diminished. Now is the time to design and implement audits and other regulatory efforts that take into greater consideration the experience of women in public life, and learn from past mistakes.

Best regards,

Kristina Wilfore
Lucina Di Meo
Co-founders, #ShePersisted