

April 11, 2025

Mark Zuckerberg  
Meta Headquarters  
1 Hacker Way  
Menlo Park, CA 94025

Dear Mr. Zuckerberg,

I am deeply disturbed and disappointed by Meta's announcement on January 7, 2025, of irresponsible changes to the content moderation policies<sup>1</sup> on its platforms that pose grave risks to the health and safety of Black communities and risk further destabilizing our republic. For nearly a decade, the NAACP Legal Defense and Educational Fund, Inc. (LDF) has invested a great deal of time and resources working with Meta as part of an informal committee advising the company on civil rights issues. Yet Meta made these content moderation policy changes without consulting or warning this group, and many of the changes directly conflict with guidance from LDF and partners. As a result, LDF can no longer in good conscience participate as part of Meta's civil rights advisory committee.

During the 2016 election, Meta (then Facebook) came under fire for failing to take action to address mis- and disinformation<sup>2</sup> and hateful content<sup>3</sup> on its platform. In 2017, LDF joined with other civil rights colleagues in sending a letter to you and then-Facebook Chief Operating Officer Sheryl Sandberg expressing our concern regarding the use of the Facebook platform to spread hateful and discriminatory content and to sow division and violence in society.<sup>4</sup> Thereafter, LDF, in partnership with a coalition of civil rights organizations, continued to push Meta to address the harmful content that continued to spread through the platform.<sup>5</sup>

Ensuring that first Facebook's and now Meta's content moderation policies limited the spread of false and hateful information was a key part of LDF's advocacy. Yet Meta has now rolled back key elements of these policies:

---

<sup>1</sup> Joel Kaplan, *More Speech Fewer Mistakes*, Meta (Jan. 7, 2025), <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>.

<sup>2</sup> Sheera Frenkel & Katie Benner, *To Stir Discord in 2016, Russians Turned Most Often to Facebook*, N.Y. TIMES (Feb. 17, 2018), <https://www.nytimes.com/2018/02/17/technology/indictment-russian-tech-facebook.html>

<sup>3</sup> Aarti Shahani, *From Hate Speech To Fake News: The Content Crisis Facing Mark Zuckerberg*, NPR (Nov. 17, 2016), <https://www.npr.org/sections/alltechconsidered/2016/11/17/495827410/from-hate-speech-to-fake-news-the-content-crisis-facing-mark-zuckerberg>.

<sup>4</sup> Letter from Civil Rights Organizations to Mark Zuckerberg (Oct. 30, 2017), [https://civilrightsdocs.info/pdf/policy/letters/2017/Letter\\_19CivilRightsGroupstoFacebook.pdf](https://civilrightsdocs.info/pdf/policy/letters/2017/Letter_19CivilRightsGroupstoFacebook.pdf).

<sup>5</sup> Press Release, Facebook's Latest Policy on Hate Speech and Voter Disinformation Misses the Mark (Jun. 26, 2020), <https://www.naacpldf.org/press-release/facebooks-latest-policy-on-hate-speech-and-voter-disinformation-misses-the-mark/>; Press Release, Readout of Civil Rights Leaders Call with Facebook (Jul. 7, 2020), <https://www.naacpldf.org/press-release/readout-of-civil-rights-leaders-call-with-facebook/>; Press Release, Facebook's Civil Rights Audit Shows Need for Reforms (Jul. 7, 2020), <https://www.naacpldf.org/press-release/facebooks-civil-rights-audit-shows-need-for-reforms/>

- Ending its third-party fact-checking system. Meta launched its factchecking program after the 2016 election in order to curb the spread of misinformation on its platforms, partnering with nonpartisan third-party entities, including news media and fact-checking nonprofits, to independently verify the accuracy of posts on their platforms.<sup>6</sup> Moving forward, instead of using third-party factcheckers, Meta will use a “Community Notes” system, whereby other users can comment on posts they believe to be false. X currently implements a similar system.<sup>7</sup> Meta will also stop limiting the distribution of fact-checked content and make it harder for users to see that posts have been flagged.<sup>8</sup>
- Allowing additional hateful speech on the platform. Meta’s policies previously limited certain speech on its platform, such as hate speech and promotion of dangerous ideologies. Meta revised those policies to clarify that it will only limit hateful conduct, not speech.<sup>9</sup> Meta also now allows certain speech that it previously prohibited, including permitting slurs based on protected characteristics; permitting users to compare protected groups to inanimate objects; and permitting allegations of mental illness or abnormality when based on gender or sexual orientation.<sup>10</sup> Finally, Meta will now wait until a user reports hateful speech to block that content or take other action, rather than searching for it using automated systems.<sup>11</sup>
- Promoting political content that is more like what users already view. Meta will start recommending political content to users based on signals such as similar content they read or liked.<sup>12</sup>

These policy changes will have concrete and predictable consequences for Black and other historically marginalized communities:

- Rise in hate speech and hate-motivated violence: These changes will allow dehumanizing hate speech to proliferate across Meta’s platforms harming Black and other communities of color, LGBTQ people, religious communities, and people with disabilities. While Meta claims to prohibit hateful conduct, this distinction is superficial where hate speech can incite hate-motivated violence.<sup>13</sup> For example, Muslim Advocates has documented how anti-Muslim hate content has led to anti-Muslim violence.<sup>14</sup> This policy change, given its

---

<sup>6</sup> Meta’s Third-Party Fact-Checking Program, Meta, <https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking> (last visited Jan. 31, 2025).

<sup>7</sup> Kaplan, *supra* note 1.

<sup>8</sup> *Id.*

<sup>9</sup> Hateful Conduct, Meta, <https://transparency.meta.com/policies/community-standards/hateful-conduct/> (last visited Jan. 31, 2025).

<sup>10</sup> *Id.*

<sup>11</sup> *Id.*

<sup>12</sup> Kaplan, *supra* note 1.

<sup>13</sup> Carlos Arcila Calderón et al., *From online hate speech to offline hate crime: the role of inflammatory language in forecasting violence against migrant and LGBT communities*, NATURE (Oc. 15, 2024), <https://www.nature.com/articles/s41599-024-03899-1> (“...online inflammatory language could be a leading indicator for detecting potential hate crimes acts...”)

<sup>14</sup> MUSLIM ADVOCATES & THE GLOBAL PROJECT AGAINST HATE AND EXTREMISM, *COMPLICIT: THE HUMAN COST OF FACEBOOK’S DISREGARD FOR MUSLIM LIFE* (2020), <https://muslimadvocates.org/wp-content/uploads/2020/10/Complicit-Report.pdf>.

timing in conjunction with a new administration, may also be viewed as an endorsement of prejudiced speech by current political leaders, which research has found emboldens audiences to declare their own prejudices and act on them accordingly.<sup>15</sup>

- **Rise in mis- and disinformation:** Meta’s policy changes increase the risk that mis- and disinformation will spread on its platforms, disproportionately harming Black people and other people of color. Mis- and disinformation threats have increased in recent years,<sup>16</sup> with targeted impacts on Black people and other communities of color.<sup>17</sup> The development and proliferation of evolving technology and AI tools have exacerbated vulnerabilities for Black communities, who have been strategically targeted by false messaging campaigns.<sup>18</sup> The tactics of disinformation campaigns have varied and the consequences have been severe—from sowing political division<sup>19</sup> to fueling mistrust in evidenced health guidance.<sup>20</sup> Importantly, a recent analysis of X’s community notes system—on which Meta’s system is modeled—concluded that the “system has fallen well short of counteracting the incentives, both political and financial, for lying, and allowing people to lie.”<sup>21</sup> The lives and safety of Black people are jeopardized by the unmitigated and targeted dissemination of mis- and disinformation, which will unmistakably increase with Meta’s abdication of its content management responsibilities.
- **Increased risk of polarization and radicalization:** Meta’s policy changes will also contribute to increased polarization in American communities. Research has

---

<sup>15</sup> Daniel Byman, *How hateful rhetoric connects to real-world violence*, BROOKINGS INST. (Apr. 9, 2021) <https://www.brookings.edu/articles/how-hateful-rhetoric-connects-to-real-world-violence/> (“Prejudiced elite speech, one study found, is particularly powerful if other elites endorse it, emboldening audiences to declare their own prejudices and act on them accordingly.”) (citing *The Trump Effect: An Experimental Investigation of the Emboldening Effect of Racially Inflammatory Elite Communication*, BRITISH J. POLIT. SCI. (Feb. 17, 2020), <https://www.cambridge.org/core/journals/british-journal-of-political-science/article/trump-effect-an-experimental-investigation-of-the-emboldening-effect-of-racially-inflammatory-elite-communication/0335108B8E4AF36CBFFA1E45816C6143#fndtn-information>)

<sup>16</sup> See, e.g., *Rise of disinformation a symptom of ‘global diseases’ undermining public trust: Bachelet*, UNITED NATIONS (Jun. 28, 2022), <https://news.un.org/en/story/2022/06/1121572>; *Disinformation is on the rise. How does it work?*, THE ECONOMIST (May 1, 2024), <https://www.economist.com/science-and-technology/2024/05/01/disinformation-is-on-the-rise-how-does-it-work>;

Jeff Cercone, *A month-by-month look at the misinformation that caught our attention in a busy news year*, POLITIFACT (Dec. 26, 2024), <https://www.politifact.com/article/2024/dec/26/a-month-by-month-look-at-the-misinformation-that-c/>.

<sup>17</sup> Whitney Tesi, *When disinformation becomes ‘racialized’*, ABC NEWS (Feb. 5, 2022), <https://abcnews.go.com/Technology/disinformation-racialized/story?id=82400863>.

<sup>18</sup> See, e.g., Matt Brown & David Klepper, *Deepfakes of Trump with Black supporters show the chilling effect of AI in election coverage*, ASSOC. PRESS (Mar. 28, 2024), <https://www.fastcompany.com/91052590/deepfakes-of-trump-with-black-supporters-shows-the-chilling-effect-of-ai-in-election-coverage>.

<sup>19</sup> See, e.g., Marquise Francis, *Influencers and popular podcasts fuel election disinformation among Black voters, report shows*, NBC NEWS (Jun. 25, 2024), <https://www.nbcnews.com/news/nbcblk/black-americans-disinformation-election-social-media-rcna157096>.

<sup>20</sup> Angela R. Pashayan, *An online pandemic: disinformation targeted at Black communities*, BULLETIN OF THE ATOMIC SCIENTISTS (Oct. 27, 2021), <https://thebulletin.org/2021/10/an-online-pandemic-disinformation-targeted-at-black-communities/>.

<sup>21</sup> Dave Lee & Carolyn Silverman, *Community Notes Can’t Save Social Media From Itself*, BLOOMBERG (Mar. 18, 2025), <https://www.bloomberg.com/opinion/features/2025-03-18/community-notes-on-x-and-meta-can-t-save-social-media-from-itself>.

demonstrated a “definitive role” that disinformation and hate speech play in polarizing societies.<sup>22</sup>

The grave impacts of Meta’s policy changes cannot be understated. Our country is in a perilous moment. Meta’s changes exacerbate the risks to the health and safety of our communities. We hope that Meta reconsiders these policy changes. At minimum, we hope that Meta will evaluate whether the Community Notes system effectively reduces user engagement with false content. We also hope that Meta publicly reports the hateful posts on its platform and publicly discloses how the policy changes impact the number of posts that are reported and removed. To this end, we hope Meta will cooperate with the various scholars collecting this data and evaluating its impact on users.

Until Meta takes action to demonstrate its commitment to civil rights, we will not participate in its civil rights advisory group. Please reach out to Puneet Cheema, Manager, Justice in Public Safety Project ([pcheema@naacpldf.org](mailto:pcheema@naacpldf.org)); Victoria Wenger, Black Voters on the Rise Counsel ([vwenger@naacpldf.org](mailto:vwenger@naacpldf.org)), and Amalea Smirniotopoulos, Senior Policy Counsel and Co-Manager of the Equal Protection Initiative, LDF ([asmirniotopoulos@naacpldf.org](mailto:asmirniotopoulos@naacpldf.org)), if you have any questions.

Sincerely,



Todd A. Cox  
Associate Director Counsel  
NAACP Legal Defense and Educational Fund, Inc.

---

<sup>22</sup> Pramukh Nanjundaswamy Vasist, et al., *The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-country Configurational Narrative*, INF. SYST. FRONT 26, 663–688 (Apr. 17, 2023) (“[T]he current study employs complexity theory and a configurational approach to investigate the impact of diverse disinformation campaigns and hate speech in polarizing societies across 177 countries through a cross-country investigation. The results demonstrate the definitive role of disinformation and hate speech in polarizing societies.”).