Title

Situated Knowledge and Algorithmic Harm in TikTok Recommendation Systems

Permalink
https://escholarship.org/uc/item/0pq4c10k

Author
Huang, Jeff

Publication Date
2026-04-30

Situated Knowledge and Algorithmic Harm in TikTok Recommendation Systems

Jeff Huang[1]

[1]Department of Computer Science, University of California, Los Angeles

*jeffhuangt@ucla.edu

Abstract: This paper uses Standpoint Theory and Theory of Change to analyze how the harms of short-form recommendation systems are produced from a dominant institutional standpoint that treats engagement metrics as neutral knowledge.

INTRODUCTION

As of 2025, TikTok has already amassed almost 2 billion active users [4]. It primarily targets a young audience between the ages of 18-29 [4]. Unlike traditional long form journalism, these short form videos compress information into 10-60 second clips that are optimized to get users' attention through shock factor or big headlines [4]. Short form content has fundamentally changed how young adults form political and cultural opinions.

Rather than simply reflecting public opinion, however, these platforms actively shape it through recommendation systems that decide what the user sees. Because short form videos are optimized for emotion and retention, they can not only influence what information is circulated, but also how the information is interpreted [5]. Most criticism of these platforms focuses on content moderation, such as which videos get removed, which creators are banned, and whether misinformation is taken down quickly enough. While these are concerning issues, they do not explain the full picture of why hate circulates so effectively on the platform [6].

This paper argues that the harms produced by short-form recommendation systems cannot be fully understood through platform metrics or moderation data alone because these systems were built from a dominant institutional standpoint that treats engagement data as neutral knowledge. Standpoint Theory is applied to show that knowledge is always socially situated and is shaped by relations of power. From this perspective, marginalized users develop analyses of the platform's operation that are structurally invisible from the dashboard view, and these analyses constitute situated knowledge in Harding's sense rather than mere lived experience. By examining experiences of South Asian and African American users, this paper shows that strong objectivity comes when analysis starts from the standpoints of those most affected by the system.

This work is in partial fulfillment of the ENGR184 course using the blueprint curriculum in Ref [1,2] and captured in a collection [3].

METHODS

This paper uses a qualitative approach grounded in science and technology studies, feminist epistemology, and media analysis. Instead of treating recommendation systems as neutral tools, this paper approaches them as sociotechnical systems shaped by corporate greed, institutional assumptions, and relations of power. This paper focuses on how TikTok defines user behavior through a series of metrics such as watch time, likes, shares, and comments [7]. These metrics determine what is recommended to the user. The data for the analysis comes from a range of data, such as public explanations of TikTok's recommendation system, moderation documents, journalism from marginalized communities, and specific case studies involving South Asian users after responding to H-1B visa discourse and Black users describing "algorithm folk theories" [8, 9].

Standpoint Theory is the primary analytical framework that this paper uses. It differentiates the perspectives of the platform and the users who are experiencing algorithmic harm. Standpoint theory is not just experience nor perspective, it is an achieved analytical position. It shows how power shapes what is deemed as legitimate knowledge. People produce it by generating counter-practices, such as 'folk theories,' and by comparing outcomes, which together create a shared account of how the platform behaves [9].

From the platform's perspective, all interaction is good interaction. A pause on a video, a share, or a 1 minute video watched all the way through can be evidence that the video is relevant or users prefer this type of content [10]. This paper questions that assumption by examining case studies where the metric is not a good representation of user experience. For example, a South Asian user searching for immigration information after the H-1B visa discourse may be treated by the platform in the same way as a user intentionally seeking racist content if both generate similar engagement signals. Likewise, a Black user warning their community about racist content may produce the same signals that another racist user looks at approvingly. Standpoint Theory makes these distinctions visible by showing how marginalized users can see harm that is unrecognized through dashboards that engineers use to track engagement.

This paper also uses concepts of situated knowledge, strong objectivity, and reflexivity to further the analysis. Situated knowledge is used to show that no account of the platform is viewless, including the accounts from the engineers, models, and the transparency reports. Strong objectivity is used to support the claim that a better understanding of how the algorithm works should start from the marginalized communities that are most affected by harm such as racial harassment. Reflexivity is the practice of examining how the position from where knowledge was produced shapes what counts as actual knowledge. Marginalized users are forced into reflexivity by friction with systems that do not accommodate them. In contrast, the platform's standpoint lacks reflexivity because its assumptions are built into the metric system itself and therefore appear more natural.

Finally, Theory of Change is used to identify the structural changes needed to be made in order to reduce these harms. The desired end state would be a short form recommendation system that does not systematically reward political polarization, harassment, or suppression of marginalized creators. Once this end state is established, the paper works to identify several key preconditions that need to be made: human readable transparency, metrics that prioritize informational diversity, and greater decision making power for communities.

RESULTS AND INTERPRETATION

TikTok is often described as just reflecting what users want to see, but in reality it does not produce a neutral picture of user behavior. Instead, it produces knowledge from a dominant institutional standpoint that treats engagement as if it were an objective measure of value. This matters because TikTok has almost 2 billion active users and is extremely influential among young people, with 43% of Americans under the age of 30 regularly getting news from TikTok [5]. The platform doesn't ask whether the content is informative, harmful, or misleading in any deep way. It asks whether the content is engaging. Engagement is then treated as knowledge about users, which reveals this is a socially situated way of knowing shaped by corporate needs.

One example is the wave of anti-Asian hate that followed Trump's announcement of the $100,000 H-1B visa [8]. In this case, the algorithm did not simply host racist content, but helped circulate and even intensify it. This day marked the highest single day use of the "j**t" slur online [8]. The platform interpreted different actions as if they were the same. South Asian Americans who were trying to search for actual immigration policy changes were also funneled into feeds that were saturated with racism and agreed with President Trump's new policy change [8]. The South Asian user searching for policy information was not engaging from a position of curiosity or agreement, but a position shaped by racial targeting. Their knowledge is situated within that social reality, while the platform's is situated within an abstract metric system that cannot tell the difference between being targeted and targeting.

This is not just mere experience. South Asians could identify something specific about how the system works. Online, "H-1B" almost always referred to South Asians, while some extremist platforms even used this term as a slur [8]. Because almost three quarters of H-1B visas go to Indians, engagement with H-1B has always involved race [11]. From the position of South Asians, what the platform said about "being interested in this topic" was actually just amplification of a racial bias against them. The claim of the system being "fair" falls apart in this lens. Amusement looks the same as agreement, harassment looks the same as friendliness, and self-defense looks like celebration. That distinction is invisible from inside the metric system but unavoidable from the position of users forced to live with its outcomes. This is where Harding's strong objectivity is useful. The dashboard isn't actually neutral. It builds in the assumption that engagement equals value, and then hides that assumption by presenting it as just a metric. Beginning analysis from the targeted user's standpoint produces a stronger account because that position must distinguish what the institutional position is structured to treat as equivalent. This is not because the targeted user's experience is more authentic, but because the dominant position has interests in not seeing what the targeted position must see.

Through repeated interaction with the platform, Black creators found that engaging with racist content drove more racist content into their feeds, that posting about social justice reduced reach, and that words like "black" in a bio could trigger moderation flags while "white supremacy" did not [12]. They developed what they call "algorithm folk theories," informal but tested accounts of how the platform suppresses content based on user identity [9]. They also developed workarounds such as, algospeak, with coded spellings like "Le$bian" to bypass automated filters, and the practice of posting unrelated content to reset how the system reads an account [13]. Each workaround is a tested claim about how the algorithm behaves. What distinguishes this from individual observation is collective epistemic labor. Black creators test claims against one another's experiences, name shared patterns, and build a community vocabulary like "algospeak," and the standpoint is achieved through that shared political work rather than by isolated noticing. This is what Standpoint Theory means by an achieved standpoint. Black creators are not just noticing the bias that engineers overlook but they have produced a more accurate working model of the system than its designers have published.

The fact that these accounts are dubbed "folk theories" rather than engineering documents is a fact about who is allowed to produce legitimate knowledge, not about whose knowledge is more accurate. This is what strong objectivity looks like in practice. This new model produced by the marginalized community's standpoint doesn't compete with the platforms' but is actually a more accurate account of how the system actually behaves. This is because the institutional position has reasons not to formalize what its users have already documented. This is also where reflexivity becomes important. Black users are forced to constantly interpret how the algorithm reads them because it simply does not fit them, while the platform does not question itself [13]. This creates an uneven relation of power. The institution determines what counts as legitimate knowledge, while marginalized users are excluded from formal knowledge production, even though they have more accurate accounts. Marginalized users' knowledge is situated, but more revealing because it exposes what the dominant account leaves out.

If TikTok's knowledge is partial like the previous sections argued, reform cannot be internal moderation because internal moderation holds the same standpoint as the institution itself. Theory of Change provides the framework for reasoning backwards from the desired end state through the preconditions required to reach it. To shift the system, there are three pre-conditions that should be met. These are often treated as parallel reforms, but the first two cannot succeed without the third, because regulators and metric designers occupy the same institutional standpoint whose limits this paper has already documented.

The EU's Digital Services Act, effective from 2024, begins to address this by requiring algorithmic transparency [14]. However, having companies release this on their own has major issues. Researchers from the University of Bath have called this information a “data abyss” [15]. Platforms like TikTok and Instagram have restricted their APIs so heavily that researchers cannot verify whether these algorithms are actually compliant with the law. Essentially, they are auditing themselves because no one else can confirm whether they are following regulations. This is a loophole where corporations remain unregulated. To make them accountable, it would require human readable documentation and external access so researchers, especially from affected communities, can verify whether these algorithms are biased. If there is no external access, their regulation only reproduces what it is supposed to check.

The second precondition is changing the algorithm's metrics, because the current metrics encode the institutional standpoint as if it were a neutral measurement. Currently, short form platforms measure success mainly through engagement [7]. These metrics reward polarization because polarizing content holds attention better, and attention is the main thing being measured. Instead of only using engagement, informational diversity would be a better strategy for measuring success. This would expose users to a larger breadth of information instead of pushing them into polarizing content and harmful speech. Platforms could also measure whether interactions increase civic engagement rather than withdrawal. If regulatory accountability is reached, then metric redesign becomes more economically reasonable. For instance, if companies face penalties and legal fees for keeping the current system, switching to an informational diversity epistemology becomes the cheaper option.

Thirdly, going back to Standpoint Theory, marginalized communities most harmed by these algorithmic biases are best positioned to take part in designing them, because the working models with the most explanatory power are currently produced outside the institution that builds the system. This is why community authority is not a third reform alongside the others but the precondition that makes the others meaningful. Without it, regulators audit using the institution's own categories and metric designers redesign using the institution's own assumptions. Organizations like Algorithmic Justice League and Data for Black Lives show how this can have real impact. For example, the Algorithmic Justice League’s Gender Shades research found that gender classification systems misclassified darker skinned women more often than lighter skinned men [16]. This research did not come from inside the companies that created the technology, but from a Black woman researcher whose own experience with facial recognition bias helped reveal the problem. Data for Black Lives similarly brings together marginalized communities to challenge algorithmic systems and push for more concrete measures beyond abstract transparency.

CONCLUSIONS

These systems are doing what their metrics define as success. They are working to maximize engagement, without regard for how much biased and harmful content is shown to users, because their metrics are incapable of measuring the distinction. Regulation must redefine accountability and metrics must value informational diversity over addictive consumption from users. Communities that are affected by the algorithms should have a voice in changing them. However, every one of these pre-conditions create less profit for companies. Knowing how these massive corporations operate, they are sure to fight against them. There has been no long lasting effort currently that fully regulates these companies. Until they are confronted with true regulations, these systems will continue to work as intended.

REFERENCES

1. Lee, Ethan, et al. "Education for a Future in Crisis: Developing a Humanities-Informed STEM Curriculum." arXiv preprint arXiv:2311.06674 (2023).

2. Sergio Carbajo, Nurturing Deeper Ways of Knowing in Science, Issues in Science & Technology, 2025, v. 41, n. 2, p. 71, doi. 10.58875/jkrw4525

3. Carbajo, S. (2026). Queered Science & Technology Center: Volume 4. Queered Science & Technology Center, 4. Retrieved from https://escholarship.org/uc/item/0zb930m0

4. Duarte, Fabio. "TikTok User Age, Gender, & Demographics (2026)." *Exploding Topics*, 16 Feb. 2026, https://explodingtopics.com/blog/TikTok-demographics

5. Tomasik, Emily, and Katerina Eva Matsa. "1 in 5 Americans Now Regularly Get News on TikTok, Up Sharply From 2020." *Pew Research Center*, 25 Sept. 2025, https://www.pewresearch.org/short-reads/2025/09/25/1-in-5-americans-now-regularly-get-news-on-TikTok-up-sharply-from-2020/

6. Gomes, André Belchior, and Aysel Sultan. "Problematizing content moderation by social media platforms and its impact on digital harm reduction." *Harm Reduction journal*, vol. 21,1 194. 9 Nov. 2024, doi:10.1186/s12954-024-01104-9, https://pmc.ncbi.nlm.nih.gov/articles/PMC11549828/

7. Zhou, Ren. "Understanding the Impact of TikTok's Recommendation Algorithm on User Engagement." *International Journal of Computer Science and Information Technology*, 2024, https://www.researchgate.net/publication/382423048_Understanding_the_Impact_of_TikTok's_Recommendation_Algorithm_on_User_Engagement

8. Stop AAPI Hate. "Keeping Count | Anti-South Asian Hate: It's Gotten Worse." *Stop AAPI Hate*, 4 Nov. 2025, https://stopaapihate.org/2025/11/04/keeping-count-anti-south-asian-hate-its-gotten-worse/

9. Karizat, Nadia, and Nazanin Andalibi. "Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance." Oct. 2021, 5. 1-44. 10.1145/3476046, https://www.researchgate.net/publication/355400840_Algorithmic_Folk_Theories_and_Identity_How_TikTok_Users_Co-Produce_Knowledge_of_Identity_and_Engage_in_Algorithmic_Resistance

10. Putri, Salsa Della Guitara, et al. "Echo Chambers and Algorithmic Bias: The Homogenization of Online Culture in a Smart Society." *SHS Web of Conferences*, v. 202, 2024, p. 05001, https://www.shs-conferences.org/articles/shsconf/pdf/2024/22/shsconf_icense2024_05001.pdf

11. U.S. Citizenship and Immigration Services. "Characteristics of H-1B Specialty Occupation Workers: Fiscal Year 2024 Annual Report to Congress." *USCIS*, 29 Apr. 2025, https://www.uscis.gov/sites/default/files/document/reports/ola_signed_h1b_characteristics_congressional_report_FY24.pdf

12. Boone, Samone. "The Erasure of Black Content Creators in Popular Media Trends: A Scoping Literature Review." *CalState*, May 2025, https://scholarworks.calstate.edu/downloads/2801pr74j

13. Bufithis, Gregory. "We Are Now Forced to Learn "Algospeak" to Avoid The Social Media Police." *Gregorybufithis*, 6 July 2022, https://www.gregorybufithis.com/2022/07/06/we-are-now-forced-to-learn-algospeak-to-avoid-the-social-media-police/

14. European Commission. "The Digital Services Act." *Shaping Europe's Digital Future*, 2026, https://digital-strategy.ec.europa.eu/en/policies/digital-services-act

15. Burnat, Florian, and Brittany Davidson. "The Accountability Paradox: How Platform API Restrictions Undermine AI Transparency Mandates." *arxiv*, 27 Mar. 2026, https://arxiv.org/html/2505.11577v3

16. Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research*, vol. 81, 2018, pp. 77–91, https://proceedings.mlr.press/v81/buolamwini18a.html?mod=article_inline