Comments to Oversight Board

Gender identity and nudity (2022-009-IG-UA & 2022-010-IG-UA)

Dear Oversight Board members,

We are writing this comment as two experts in technology policy, content moderation, social media, and gender. Oliver Haimson is an assistant professor at University of Michigan School of Information and the Institute for Research on Women and Gender, and a recipient of a National Science Foundation award focused on how to make social media content moderation more equitable. He has published research about social media content moderation at top venues in computing and in communication and media studies, including the ACM Conference on Computer-Supported Cooperative Work (CSCW) and *New Media & Society* (Haimson et al., 2021; Thach et al., 2022). Kendra Albert is a technology lawyer at Harvard Law School's Cyberlaw Clinic, as well as a lecturer in Women, Gender, and Sexuality. Their research focuses on the intersection of computing, gender, and society, and has been published in top venues including law journals and the ACM Conference on Fairness, Accountability, and Transparency (FAccT) (Albert et al., 2020; Albert & Delano, 2021). In addition to our professional expertise, we are also trans people ourselves, which makes us intimately familiar with the type of marginalization Meta imposes on transgender people.

The case before the Oversight Board is an archetypal example of the harm that transgender people attempting to use social media face. In it, two trans people did not violate site policies yet were harmed not only by their images and speech being silenced, but also by the site limiting their ability to crowdfund money for surgery. It is not an isolated incident. In aggregate, both "neutral" policies that are disproportionately enforced against trans users and policies that target trans users have the effect of limiting trans people's abilities to post content related to their marginalized identity, thus silencing their participation in the public sphere. Issues for transgender users on online platforms cannot be solved by magical thinking around equitable enforcement; they require a deeper reimagination of how bodies are understood and regulated.

Below, we discuss examples and evidence of how these processes play out in practice, and what the Oversight Board can learn from the examples in crafting a decision on the case before it. In doing so, it may be impossible for an organization like Meta to sufficiently limit the cisgender gaze in its own content moderation policies and enforcement to equitably treat transgender users. Fully addressing trans marginalization on Meta may necessitate changes to underlying nudity, sexual

solicitiation, and other policies to eliminate the engines of disproportionate harm, rather than attempting to create exceptions for transgender users and their content. But it is not just the policies that are problematic here—the enforcement of these policies must be carried out in a way that does not further marginalize trans people on the platform.

Content Moderation's Impact on Transgender People

Research has shown that social media content moderation disproportionately harms transgender people in practice (Dinar, 2021; Haimson et al., 2021; Haimson & Hoffmann, 2016; Smith et al., 2021). Haimson et al.'s recent paper found that transgender social media users were significantly more likely to experience both content removals and account removals on social media sites including Instagram and Facebook, even in instances when they did not violate the site's policies or guidelines (Haimson et al., 2021). For instance, content removed as "adult" despite following site guidelines was one of the major categories of content removed more often from transgender users - this type of content was removed from 21.2% of participants in the study, a rate roughly five times higher than cisgender participants (Haimson et al., 2021). Disproportionate removals of "adult" content does not mean that trans people are posting nude content more often than other users - it means that their content, whether or not it does actually depict nudity or sexual content, is flagged as such by social media site's algorithms or human moderators. Further, even when trans users appeal these faulty decisions, removal decisions are unlikely to be overturned (Smith et al., 2021).

Computational content moderation often does not work in "edge" cases, such as when people from marginalized groups post content on social media (Gillespie, 2018). For instance, to an algorithm, breasts in the context of breastfeeding, which is allowed on Facebook, are often indistinguishable from breasts in contexts that are forbidden on Facebook. Similarly, trans bodies may throw a wrench in algorithmic content moderation and may appear computationally similar to forbidden content. Further, trans bodies are likely vastly underrepresented in machine learning training sets, and when they are included, are likely often misgendered (Scheuerman et al., 2019).

Thus, we cannot rely on computational content moderation to make decisions about trans bodies on social media. And yet we also cannot rely on human moderators, the vast majority of whom are cisgender, to make decisions about trans bodies. Based on the fact that the photos in question were appealed twice and still removed, human moderators likely were following a rule book that did not properly account for trans bodies and lives. Content moderation decisions about trans content on social media

requires extra care and understanding that algorithms and cisgender moderators, neither of whom have been properly trained about transness, have not exercised.

Meta's sexual solicitation policy is in itself harmful. But Meta's sexual solictation policy is not even the issue here, as crowdfunding for medical care is not comparable to sexual solicitation. Crowdfunding and other forms of mutual aid are essential to trans communities (Barcelos, 2022; Farnel, 2014; Fritz & Gonzales, 2018; Gonzales & Fritz, 2017; Spade, 2020), as trans people not only earn significantly less money than cisgender people (James et al., 2016) but also have additional medical expenses that are often not covered by insurance. Crowdfunding is not the same thing as sexual solicitation, and it matters deeply that Meta make a distinction here. Trans people must crowdfund money to cover surgery, which is a form of healthcare that may literally save trans people's lives (Bauer et al., 2015), and social media are an important mechanism that many trans people use to garner support for their crowdfunding campaigns. Given these circumstances, removing crowdfunding content and labeling it as sexual solicitation seems especially cruel and unwarranted.

The Context of Meta's Regulation of Transgender Bodies

Although it may be tempting to understand the case before the Oversight Board as a one-off failure of either Instagram's human content moderation teams or its algorithmic systems, transgender people find themselves often accused for sexual solicitaiton for any form of nudity, hint of sexuality, or even just for existing.

Even outside of the space of content moderation, transgender bodies are often seen as inherantly sexual by cisgender people. This dynamic occurs even when the behavior or speech that trans people are engaged in is not sexual. As an example from physical space, laws that prohibit "loitering for the purpose of prostitution," such as the New York law that was referred to as the "Walking While Trans" ban, are disproportionately enforced against transgender women of color, whether said women were engaged in sex work or not (Arnold 2020, Diaz 2021). Such views may have their root in transmisogynistic concepts such as autogynephilia, combined with the inherent racism of policing systems (Serano 2007). Meta's sexual solicitation policy seems to be functioning in a similar manner here.

Likewise, cisgender system designers often project their own notions of binary sex/gender concepts onto trans people, with tragicomic results. Facebook's policies regarding "female nipples" are an example of this, as they create a category that is based in neither biology nor gender, as projects like Genderless Nipples have pointed out (Toor 2016). Although said policy is not at issue here, as the users in questions had covered their nipples, the creation of policies that force users to attempt to identify

whether their nipples are too "female" could be dysphoria-inducing. Again, there are parallels to offline contexts, like systems of medical gatekeeping, where transgender people are often required to convinced cisgender doctors that they are trans enough (Stryker 2017).

The inherent sexualization of trans bodies and behaviors, and the cis-centrism of policies used to categorize transgender bodies have profound negative impacts on transgender people. They consistently remind transgender people that they are not welcome in public, putting pressure on them to "pass" (i.e., be read as cisgender) and to wear clothing, take transition steps, or otherwise behave in ways that are less likely to attract notice. And they put disproportionate pressure on transgender creators and users who wish to discuss their own transgender status or raise awareness about issues that disproportionately affect transgender people. Even when trans people comply with the rules, they still risk content deletion and bans, both at the hands of automated systems and from human reviewers.

The Oversight Board

In light of this background and how transgender people are affected by content moderation, the Oversight Board should make the following recommendations to Meta:

- The policies that currently exist around sexual solicitation and nudity do not respect the rights of transgender and nonbinary people. To do so, they will need to be reevaluated.
- Meta's nudity and sexual solicition policies should be redeveloped, centering transgender and nonbinary people, including sex workers, with transgender people as active leaders and constituents in developing the policies, as opposed to tokens or tacked-on afterthoughts.
- The Oversight Board should communicate to Meta that persistent enforcement errors that disproportionately affect particular communities are, in themselves, a sign of system failure. If the enforcement was in error, it should reinforce the need to reconsider the policies applied to transgender bodies and users, as opposed to obviating it.

Respectfully submitted,

Kendra Albert
Clinical Instructor / Cyberlaw Clinic, Harvard Law School
Lecturer / Program on Women, Gender, and Sexuality, Harvard University

Oliver L. Haimson
Assistant Professor of Information / School of Information, University of Michigan

References

Albert, K., Armbruster, E., Brundige, E., Denning, E., Kim, K., Lee, L., Ruff, L., Simon, K., & Yang, Y. (2020). FOSTA in Legal Context, Columbia Human Rights Law Review.

Albert, K., & Delano, M. (2021). This Whole Thing Smacks of Gender: Algorithmic Exclusion in Bioimpedance-based Body Composition Analysis. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 342–352. https://doi.org/10.1145/3442188.3445898

Arnold, A. (2020, July 22). *Police Can Arbitrarily Arrest Trans Women Under This Law*. The Cut. https://www.thecut.com/2020/07/walking-while-trans-law-in-new-york-explained.html

Barcelos, C. (2022). The Affective Politics of Care in Trans Crowdfunding. *TSQ: Transgender Studies Quarterly*, *9*(1), 28–43. https://doi.org/10.1215/23289252-9475495

Bauer, G. R., Scheim, A. I., Pyne, J., Travers, R., & Hammond, R. (2015). Intervenable factors associated with suicide risk in transgender persons: A respondent driven sampling study in Ontario, Canada. *BMC Public Health*, *15*(1), 525. https://doi.org/10.1186/s12889-015-1867-2

Diaz, J. (2021, February 3). New York Repeals "Walking While Trans" Law. *NPR*. https://www.npr.org/2021/02/03/963513022/new-york-repeals-walking-while-trans-law

Dinar, C. (2021). *The state of content moderation for the LGBTIQA+ community and the role of the EU Digital Services Act* (p. 23). Heinrich-Böll-Stiftung.

Farnel, M. (2014). Kickstarting trans*: The crowdfunding of gender/sexual reassignment surgeries. *New Media & Society*, *17*(2), 215–230. https://doi.org/10.1177/1461444814558911

Fritz, N., & Gonzales, A. (2018). Not the Normal Trans Story: Negotiating Trans Narratives While Crowdfunding at the Margins. *International Journal of Communication*, *12*(0), 20.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.

Gonzales, A., & Fritz, N. (2017). Prioritizing Flexibility and Intangibles: Medical Crowdfunding for Stigmatized Individuals. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2371–2375. https://doi.org/10.1145/3025453.3025647

Haimson, O. L., Delmonaco, D., Nie, P., & Wegner, A. (2021). Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 466:1-466:35. https://doi.org/10.1145/3479610

Haimson, O. L., & Hoffmann, A. L. (2016). Constructing and enforcing "authentic" identity online: Facebook, real names, and non-normative identities. *First Monday*, *21*(6). https://doi.org/10.5210/fm.v21i6.6791

James, S. E., Herman, J. L., Rankin, S., Keisling, M., Mottet, L., & Anafi, M. (2016). *The Report of the 2015 U.S. Transgender Survey*. National Center for Transgender Equality. http://www.transequality.org/sites/default/files/docs/USTS-Full-Report-FINAL.PDF

Scheuerman, M. K., Paul, J. M., & Brubaker, J. R. (2019). How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis and Image Labeling Services. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 33.

Serano, J. (2007). *Whipping Girl: A Transsexual Woman on Sexism and the Scapegoating of Femininity*. Seal Press.

Smith, S. L., Haimson, O. L., Fitzsimmons, C., & Brown, N. E. (2021). *Censorship of Marginalized Communities on Instagram*. Salty.

Spade, D. (2020). *Mutual Aid: Building Solidarity During This Crisis (and the Next)*. Verso Books.

Stryker, S (2017). Transgender History: The Roots of Today's Revolution, 2nd Edition. Seal Press.

Thach, H., Mayworm, S., Delmonaco, D., & Haimson, O. (2022). (In)visible moderation: A digital ethnography of marginalized users and content moderation on Twitch and Reddit. *New Media & Society*, 146144482211098. https://doi.org/10.1177/14614448221109804

Toor, A. (2016, December 6). *Genderless Nipples exposes Instagram's double standard on nudity*. The Verge. https://www.theverge.com/2016/12/6/13852900/genderless-nipples-instagram-censorship-policy