How Instagram's algorithm is censoring women and vulnerable users but helping online abusers

Carolina Are

To cite this article: Carolina Are (2020): How Instagram's algorithm is censoring women and vulnerable users but helping online abusers, Feminist Media Studies, DOI: 10.1080/14680777.2020.1783805

To link to this article: https://doi.org/10.1080/14680777.2020.1783805

How Instagram's algorithm is censoring women and vulnerable users but helping online abusers

Carolina Are

City, University of London, London, UK

Initially, social media platforms seemed to aid activists and their cause, becoming an opportunity for change in the face of institutional failure (Sonja Vivienne 2016). Social media movements such as the Arab Spring, #MeToo, #BlackLivesMatter and #OccupyWallStreet painted these platforms as heralding freedom of expression (Anabel Quan-Haase and Luke Sloan 2017; Smith et al. 2014). The founder of Facebook (and owner of Instagram) Mark Zuckerberg (2018) views technology as a "democratizing force for putting power in people's hands," and for a while, this hope felt tangible. Yet, in 2020, social networks' own infrastructure actually often harms and censors their most vulner- able users. Through a dual focus on my PhD project and activism, this article will draw from research on online abuse and experiences of and protests against Instagram's algorithm bias. It will conclude by providing both user and governance recommendations for better moderation on the platform.

Offline vulnerabilities have followed women online. Even celebrities are being drawn off social media by repetitions of offline gendered and racial hatred: actor Leslie Jones, targeted via endless misogynist and racist comments for being a black woman in the *Ghostbusters* remake through Twitter, ended up deleting her account (Caitlin E Lawson 2018). Cyberflashing—sharing nude pictures via Bluetooth in public spaces (Jay Boulos 2019)—and, similarly, online harassment and unsolicited "dick pics" (Sophia Ankel 2018; Emma Jane 2015) are examples of how social media present what Alison Harvey 2019 calls "aggressive architecture" which sees platforms' react with "active inactivity" instead of protecting their users. Harassment has emotional, psychological and economic costs for victims, making women stop contributing to online spaces and cutting them off from work and/or public life. The same platforms that were going to give them voice are also giving users new opportunities to harass, insult and silence them.

Gendered policing is a double-edged phenomenon. In addition to harassment, women's bodies, nudity, sex and sexuality also bear the brunt of social media's algorith- mic censorship, replicating the male gaze online (K Jarrett, B Light and S Paasonen 2019). Such a power imbalance within social media has become even more apparent following the approval of FOSTA/SESTA, an exception to Section

230 of the Communication Decency Act in the United States that ruled platforms were not liable for what was posted on them (115th Congress, 2017-2018). As a result, social media companies have been deleting and censoring an increasing number of posts showing skin for fear of being seen as a utility that "promotes or facilitates prostitution," (115th Congress, 2017–2018).

Following FOSTA/SESTA, a variety of sex workers have had their accounts deleted, while communities ranging from Carnival dancers to athletes had their posts deleted or hidden by Instagram (Carolina Are 2019a; Sharine Taylor 2019).

As a pole dance performer and instructor, I have seen many of my posts censored or affected by Instagram's "shadowban," a form of light censorship targeting "vaguely inappropriate content" that is hidden from the platform's explore page (John Constine 2019), preventing me, like many in my industry, from reaching wider audiences and finding more work. When in July 2019 a petition signed by nearly 20,000 pole dancers forced Instagram to officially apologise to pole dancers through my blog, it became clear that only an uproar would force the platform to even admit any form of wrongdoing. In the apology, the platform denied wanting to target specific communities, arguing content and hashtags were moderated "in error" and claiming that, due to the exponential amounts of content posted on Instagram every day, the platform is subject to making mistakes (Carolina Are 2019b). Yet, mistakes or not, Instagram's infrastructure seems to favour certain users as opposed to others.

As one of the founding members of EveryBODYVisible, a campaign against Instagram censorship launched in October 2019 and counting burlesque performer Dita Von Teese amongst its supporters, I have been part of informing our audience and experimenting with different techniques to boost their visibility online. From seeing women's engage- ment increase when they changed gender to male on the platform, to tagging the chiefs of Facebook and Instagram during our campaign, so that their "tagged" section on their profile ended up featuring a variety of women's bottoms, we made so much noise that Instagram CEO Adam Mosseri had to recognise our demands were reasonable in a story post (Carolina Are 2019c).

At present, fighting for more equal moderation is left to activists and those affected by censorship or harassment. As a PhD student focusing on online harassment, I have witnessed hateful comments and the lack of moderation surrounding them driving women off platforms and having to deal with their traumatic consequences without support. As a pole dancer, I have been censored multiple times. As a blogger, I have interviewed Instagram's press team multiple times, asking for clarifications on their algorithmic moderation without receiving more than recycled press release information as an answer.

As a result of these experiences, I hope to share recommendations for both censored users and for the wider governance of social media. These recommendations are informed by a variety of data gathered through autoethnography, including my own experience with Instagram censorship of pole dancing and women, interviews with Instagram's press team and experiences from an ongoing PhD research on online abuse.

Users looking to support people targeted by social media censorship and/or harassment should:

- ● Constantly check accounts they support that have claimed to be vulnerable to censorship/harassment;
- ● Constantly call out unfair moderation practices, whether for responses to harass- ment or for censorship;
- ● Demand better moderation, e.g. through platforms' "Help" or "Report a problem" functions.

Yet, social media moderation is as much related to expression as it is political—and governments should help in this fight. Future changes in social media governance should therefore consider:

- ● Actively breaking up the monopolies of social media giants such as Facebook and Google (David Kaye 2019);
- ● Pressing for clarity and transparency about moderation techniques and decisions about user content and data (ibid);
- ● Pressing social media giants for the implementation of international human rights standards into social media content moderation, considering that, in respect of Article 10 of the European Convention of Human Rights, publishing content that is shocking or offensive (within reason) is not a crime (Case of Oberschlick (no. 2) v. Austria 1997; Kaye 2019; European Court of Human Rights, 1950).

It is not sustainable for large parts of their user populations to continue being silenced by and targeted on social media: if social media architecture is kept as it is, offline inequalities may become even greater online, and the value that social media platforms could provide to our society will be lost. More equal moderation is a win-win situation: it justifies platforms' existence and it improves users' experience.

Social media platforms have become a form of civic space. Because of this, platforms need to be held accountable about their biases, and they need to be more transparent about the rules that govern them.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Carolina Are http://orcid.org/0000-0003-1110-3155

1. **References**

115th Congress. 2017-2018. "H.R.1865 - Allow States and Victims to Fight Online Sex Trafficking Act of 2017." congress.gov. https://www.congress.gov/bill/115th-congress/house-bill/1865

Ankel, Sophia. 2018. “Unsolicited ‘Dick Pics’ Shouldn’t Be Laughed Off, but Criminalised.” The Guardian. https://www.theguardian.com/commentisfree/2018/oct/24/dick-pics-criminalised-flashing-intimate-photo-consent

Are, Carolina. 2019a. “Instagram Apologises to Pole Dancers about the Shadowban.” Bloggeronpole. com. https://bloggeronpole.com/2019/07/instagram-apologises-to-pole-dancers-about-the- shadowban/

Are, Carolina. 2019b. “Instagram Denies Censorship of Pole Dancers and Sex Workers.” Bloggeronpole.com. https://bloggeronpole.com/2019/07/instagram-denies-censorship-of-pole- dancers-and-sex-workers/

Are, Carolina. 2019c. “Instagram Censors EveryBODYVisible Campaign against Instagram Censorship – LOL.” Bloggeronpole.com. https://bloggeronpole.com/2019/10/everybodyvisible/

Boulos, Jay. 2019. “Cyber-flashing: ‘I Froze When Penis Picture Dropped on to My Phone’” BBC. https://www.bbc.co.uk/news/uk-48054893

Case of Oberschlick (no. 2) v. Austria. 1997. *The International Journal of Human Rights* 1 (3): 89–91. doi:10.1080/13642989708406687.

Constine, John. 2019. “Instagram Now Demotes Vaguely ‘Inappropriate’ Content.” TechCrunch. https://techcrunch.com/2019/04/10/instagram-borderline/

Jane, Emma. 2015. “Flaming? What Flaming? The Pitfalls and Potentials of Researching Online Hostility.” *Ethics Information Technology* 17: 65–87. doi:10.1007/s10676-015-9362-0.

European Court of Human Rights. 1950. “European Convention on Human Rights.” https://www. echr.coe.int/Documents/Convention_ENG.pdf

Harvey, Alison. 2019. “Tits or GTFO: The Aggressive Architecture of the Internet.” Flowjournal.org. http://www.flowjournal.org/2019/05/tits-or-gtfo-the-aggressive-architecture-of-the-internet- alison-harvey-university-of-leicester/#identifier_2_37691

Kaye, David. 2019. *Speech Police - The Global Struggle To Govern The Internet*. New York: Columbia Global Reports.

Lawson, Caitlin E. 2018. “Platform Vulnerabilities: Harassment and Misogynoir in the Digital Attack on Leslie Jones.” *Information, Communication & Society* 21 (6): 818–833. doi:10.1080/ 1369118X.2018.1437203.

Paasonen, S, K Jarrett, and B. Light. 2019. *#NSFW: Sex, Humor, And Risk In Social Media*. Cambridge: MIT Press.

Sloan, Luke, and Anabel. Quan-Haase. 2017. *The SAGE Handbook of Social Media Research Methods*. London: SAGE Publications.

Smith, Marc A. Rainie, Lee, Shneiderman, Ben and Himelboim, Itai. 2014. “Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters.” Report, Washington: Pew Research Center, pp.1–56. Available at: https://www.pewresearch.org/internet/2014/02/20/mapping-twit- ter-topic-networks-from-polarized-crowds-to-community-clusters/

Taylor, Sharine. 2019. “Instagram Apologises For Blocking Caribbean Carnival Content.” Vice. https:// www.vice.com/en_in/article/7xg5dd/instagram-apologises-for-blocking-caribbean-carnival- content

Vivienne, Sonja. 2016. *Digital Identity and Everyday Activism - Sharing Private Stories with Networked Publics*. Basingstoke: Palgrave MacMillan.

Zuckerberg, Mark 2018. “A Blueprint for Content Governance and Enforcement.” Facebook. Available at: https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-govern ance-and-enforcement/10156443129621634/.