



Selected Papers of #AoIR2022:  
The 23rd Annual Conference of the  
Association of Internet Researchers  
Dublin, Ireland / 2-5 Nov 2022

## EXPLOITATION IN ONLINE CONTENT MODERATION

Karen Frost-Arnold  
Hobart & William Smith Colleges

This paper presents a normative framework for evaluating the moral and epistemic exploitation that online content moderation workers experience while working for social media companies (often as subcontractors or as Mechanical Turk workers). I argue that the current labor model of commercial content moderation (CCM) is exploitative in ways that inflict profound moral harm and epistemic injustices on the workers. This detailed account of exploitation enables us to see more clearly the contours and causes of the moral and epistemic injustice involved in CCM, and helps us understand precisely why these forms of exploitation are unjust. It also suggests some practical solutions for a more just labor model for the moderation work that shapes our online ecosystem.

To show that CCM is *exploitative in a moral sense*, I draw on Ruth Sample's account of exploitation (Sample 2003). Sample defines exploitation as "interacting with another being for the sake of advantage in a way that fails to respect the inherent value in that being" (Sample 2003, 57). To avoid exploiting someone, we must "take their needs into account" (Sample 2003, 81). This has failed to happen in the development of the labor model of CCM; moderators have basic human needs that have not been taken into account. Thus, they have been treated with disrespect. To make this case, I argue that CCM violates workers' basic needs for emotional health and affiliation (i.e., relationships) (cf. Nussbaum 1999, 41). The emotional trauma of PTSD reported by CCM workers and the emotionally desensitizing nature of the work both harm the mental health of workers. The intense exposure to images of child sexual exploitation, violence, and other forms of human cruelty does emotional harm that frequently damages moderators' relationships (Roberts 2016; 2019; Chen 2017; Bowden 2017; LaPlante 2017; Solon 2017). Additionally, non-disclosure agreements prevent workers from discussing their difficult work with friends and family—thus interfering with workers' basic needs for connection and affiliation (Roberts 2016).

To show that CCM is *epistemically exploitative*, I develop an account of a novel form of epistemic injustice, which I call "epistemic dumping." The philosophical literature on epistemic injustice reveals the many ways in which unjust social systems and practices disrespect marginalized people in their capacity as knowers (Fricker 2007; Dotson 2014). I define *epistemic dumping* as a form of epistemic injustice in which the labor of

Suggested Citation (APA): Frost-Arnold, K. (2022, November). *Exploitation in Online Content Moderation*. Paper presented at AoIR 2022: The 23rd Annual Conference of the Association of Internet Researchers. Dublin, Ireland: AoIR. Retrieved from <http://spir.aoir.org>.

sorting, handling, and disposing of the toxic epistemic trash of epistemic communities is disproportionately performed by some (typically traditionally oppressed) groups, while other (typically traditionally dominant) groups benefit from this labor, though usually being oblivious to its performance or dismissive of the epistemic labor involved in handling the trash. My concept of *epistemic trash* draws on the work of Sarah Roberts and Sarah Jeong. Roberts uses the language of “digital detritus” and “digital refuse” to refer to the harmful online content which moderators are charged with handling (Roberts 2016). Jeong also uses the metaphor of garbage, using it to refer to “undesirable content” ranging from spam to online harassment (Jeong 2015, 11). Thinking about the content that moderators remove as a form of epistemic trash helps us recognize the epistemic labor of CCM and also ask questions about who “takes out” the trash and how those workers are treated. I argue that the labor model of CCM exemplifies epistemic dumping because the labor of removing online trash is disproportionately performed by low-wage subcontractors (often in the Global South) whose epistemic skills are disrespected. Social media companies have chosen to treat their labor as “low-skill” work that can be performed in an assembly-line format, paid poorly, and hidden from sight (Roberts 2016; 2019; Block and Riesewieck 2018). Moderators are often treated as human computers, rather than as epistemic agents. This is a kind of epistemic objectification that disrespects them as knowers. For example, Punsmann, a former Facebook moderator, writes, “The mental operations [of moderation] are evaluated as being too complex for algorithms. Nevertheless moderators are expected to act as a computer. The search for uniformity and standardization, together with the strict productivity metrics, lets not much space for human judgment and intuition” (Punsmann 2018). Moderators are epistemically objectified and their human judgment is not valued. Thus, CCM is epistemically exploitative because it disproportionately dumps the toxic epistemic content of our online epistemic communities on moderators while simultaneously disrespecting their epistemic skills at handling digital detritus.

This account of the epistemic injustice involved in CCM helps us imagine another world in which the difficult work moderators perform is valued as skilled and nuanced. Together, the ethical and epistemic accounts of exploitation support normative arguments for better compensation, increased mental health support, more respect, and more recognition for the laborers who perform the difficult labor of tidying up our online environments.

## **Methodology**

This is a project in ethics and epistemology of the internet. While the theoretical tools are primarily philosophical in nature, the paper also draws on internet studies scholarship on the nature of online labor and content moderation (e.g. Gray and Suri 2019; Roberts 2019). To make it accessible to an interdisciplinary audience, the paper will begin with an introduction to the literature on the ethics of exploitation and epistemic injustice, as well as providing an overview of the landscape of CCM.

## **References**

Block, Hans, and Moritz Riesewieck. 2018. *The Cleaners (Im Schatten Der Netzwelt)*. Gebrueder Beetz Filmproduktion.

- Bowden, Rasalyn. 2017. "CCM Workers: Tales from the Moderation Screen." Presented at the All Things in Moderation Conference, University of California, Los Angeles, December 7. <https://atm-ucla2017.net/>.
- Chen, Adrian. 2017. "The Human Toll of Protecting the Internet from the Worst of Humanity." *The New Yorker*, January 28, 2017. <https://www.newyorker.com/tech/elements/the-human-toll-of-protecting-the-internet-from-the-worst-of-humanity>.
- Dotson, Kristie. 2014. "Conceptualizing Epistemic Oppression." *Social Epistemology* 28 (2): 115–38.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and Ethics in Knowing*. New York: Oxford University Press.
- Gray, Mary L., and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. New York: Houghton Mifflin Harcourt.
- Jeong, Sarah. 2015. *The Internet of Garbage*. Jersey City, NJ: Forbes.
- LaPlante, Rochelle. 2017. "CCM Workers: Tales from the Moderation Screen." Presented at the All Things in Moderation Conference, University of California, Los Angeles, December 7. <https://atm-ucla2017.net/>.
- Nussbaum, Martha Craven. 1999. *Sex & Social Justice*. New York: Oxford University Press.
- Punsmann, Burcu Gültekin. 2018. "Three Months in Hell: What I learned from three months of Content Moderation for Facebook in Berlin." *SZ Magazin*. January 6, 2018. <https://sz-magazin.sueddeutsche.de/internet/three-months-in-hell-84381>.
- Roberts, Sarah T. 2016. "Digital Refuse: Canadian Garbage, Commercial Content Moderation and the Global Circulation of Social Media's Waste." *Wi: Journal of Mobile Media* 10 (1): 1–18.
- . 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven: Yale University Press.
- Sample, Ruth J. 2003. *Exploitation: What It Is and Why It's Wrong*. Lanham, MD: Rowman & Littlefield.
- Solon, Olivia. 2017. "Underpaid and Overburdened: The Life of a Facebook Moderator." *The Guardian*, May 25, 2017. <http://www.theguardian.com/news/2017/may/25/facebook-moderator-underpaid-overburdened-extreme-content>.