A Keyword Entry on “Commercial Content Moderators”

Alyssa Miranda

Abstract

In today’s digital economy, our enthusiasm for social media platforms often causes us to overlook a hidden network of labourers who endeavour to sustain friendly online spaces where people are free to creatively express themselves and feel safe from hostile behaviour. These individuals, referred to as “commercial content moderators,” play a crucial role in balancing the diverse interests of site visitors, business corporations, and content creators. Commercial content moderators must regularly endure undesirable working conditions and psychological stress, which can take a heavy toll on the quality of their personal lives and raises ethical concerns regarding labour regulations in the countries where they are employed. This paper will examine the role of commercial content moderators in a scholarly, policy, and public-related context. It will also identify experts, policy events, and key resources pertaining to the invisible pool of knowledge workers in this industry.

Keywords: Commercial content moderation, precarity in digital labour, online harassment, human rights

Introduction

Keyword studies were pioneered by University of Cambridge academic Raymond Williams, who became fascinated by terminologies that had different contextual meanings depending on the culture and time period in which they were used, after serving abroad in World War II (Raymond, 1985, p. 11-12). Today, keywords are increasingly becoming an important staple of the information society (Peters, 2016, p. xiii). Not only do they seek to define and explain the relevance of prevalent digital phenomena, but they also contain a collection of resources that serve as a point of reference for further reading among scholars and individuals who are interested in the topic (Peters, 2016, p. xiv-xv). This keyword entry will examine the usage of the term “commercial content moderators” in a scholarly,
policy, and public related context. In particular, it will investigate and describe how commercial content moderators are instructed to discretionally use their services in a discriminatory manner for profit, exploitation and precarity in digital labour, and the tensions that arise when balancing users’ rights to freedom of expression with their equal rights to protection from online harassment.

**The Employment of Discriminatory Practices for Profit**

In scholarly discourse, commercial content moderators are viewed as “digital gatekeepers” for social media platforms and websites (Roberts, 2016b, p. 148). They play an important role in protecting the reputation of company brands by deleting inappropriate material in user-generated content that may be posted by site visitors or members (Roberts, 2016b, p. 147). The two categories of content moderation utilized by most social media companies are active moderation and reactive moderation (Chen, 2014). “Active moderation” involves the real-time screening of every post before it is uploaded to the social media platform, which is a very time-consuming and laborious task (Chen, 2014). During this process, commercial content moderators must frequently refer to a list of guidelines that include keywords such as gore, pornography, and racism, to remind themselves of what they are looking for in moments of uncertainty (Chen, 2014). By contrast, “reactive moderation” only requires commercial content moderators to filter posts that have been flagged as obscene by users based on site regulations, local norms, and values (Chen, 2014; Roberts, 2016b, p. 147). The rigorous methods and proactive reasoning behind active moderation have made it more common among social media platforms like Facebook and YouTube, in comparison to reactive moderation (Chen, 2014).

During the decision-making process of whether or not to delete ambiguously offensive user-generated content, moderators are required to consider the potential virality of these posts (Roberts, 2016b, p. 150). The term “virality” refers to content that has accumulated millions of views, comments, likes, shares, and parodies by fans as a result of its online popularity (Roberts, 2016b, p. 150). If the likelihood of a questionable post going viral is high, commercial content moderators are instructed to allow it to be uploaded onto the company’s website or platform because this will increase the brand’s recognition in mainstream media, and thus boost their revenue (Roberts, 2016b, p. 150). However, these business-motivated decisions are often made at the expense of vulnerable minority groups, as many posts that go viral contain humour that is racist, sexist, classist, or homophobic (Roberts, 2016b, p. 150). For example, a news report of Antoine Dodson, the brother of a rape victim, went viral on YouTube due to the popularly perceived humour found in his accent, flamboyant mannerisms, and use of slang vocabulary, even though it has offended many African-Americans, the LGBTQ community, and members of the working class (Roberts, 2016b, p. 153). This raises many ethical concerns surrounding the use of profit-driven practices that continue to hide their discriminatory effects against minority groups, since very few people are aware of the work that moderators do (Roberts, 2016b, p. 156).
Exploitation and Precarity in Digital Labour

In a policy-related context, active moderation services for social media websites and platforms are usually outsourced to countries like the Philippines due to soft labour laws and a lack of enforcement mechanisms for legal regulations (Chen, 2014). Commercial content moderators from developing countries are often paid poorly and work long hours in spite of the important and specialized work that they are responsible for doing (Chen, 2014). Moderators must familiarize themselves with the culture that the majority of their audience identifies with in order for them to correctly interpret the intentions behind each post within a matter of seconds (Chen, 2014).

Furthermore, the disturbing text and imagery containing pornographic, violent, and/or morally objectionable content that commercial content moderators encounter on a daily basis can have a severe psychological effect on their wellbeing (Chen, 2014). Moderators reportedly experience symptoms such as an overactive sex drive, or an irrational paranoia towards other people, which heavily interferes with their personal and professional lives by negatively affecting their interpersonal relationships with partners and colleagues (Chen, 2014). In an interview with Wired magazine, Jane Stevenson, the founder of an occupational health and wellbeing service called Workplace Wellbeing, has equated these adverse effects with those experienced by child pornography and anti-terrorism investigators from the law enforcement profession who are suffering from post-traumatic stress disorder (Chen, 2014). Stevenson explains that although the police department provides its agents with extensive services to mitigate the emotional damage caused in their line of work, companies either offer inadequate psychiatric support to the moderators they employ, or none at all (Chen, 2014).

Moderators are also unable to discuss the psychological stress of their work with their family or friends due to strict non-disclosure agreements that they sign when they start their contracts, and often cannot afford the services of external psychologists (Roberts, 2016a). Low wages, long hours and an inability to confide in trusted individuals often results in commercial content moderators choosing to leave their positions to seek other work that is less onerous (Chen, 2014). Nonetheless, the industry itself remains highly esteemed, given the affiliation of prestigious brands (Chen, 2014). Policies and regulations need to be implemented to address issues of abuse and negligence towards moderators that are prevalent within this industry.

Freedom of Expression and Online Harassment

Public dialogue about commercial content moderators focuses on their capacity to balance users’ rights to freedom of expression with other users’ rights to protection from hate speech and virtual harassment (Roberts, 2016b, p. 152). The Internet is increasingly becoming a site for activism and for awareness of ongoing social issues (Roberts, 2016a). Consequently, moderators are put under a lot of pressure to make the right decisions when it comes to such matters, which is why it is imperative for them to be well-versed in the norms and values of the societies that have access to the content that they are evaluating (Chen, 2014). If they incorrectly categorize a social activist’s post as offensive or
fail to remove an obscene comment directed at another user, then their client company will receive significant backlash from consumers and media outlets, which is damaging to their reputation and societal influence. For example, feminists continue to express outrage and retaliate against Facebook’s ban on photographs of mothers breastfeeding their children because it violates the website’s community guidelines governing nudity, which has attracted a large amount of negative publicity (Gibbs, 2016). Although the decisions they make are subject to much critique, moderators play an important public role by mediating conflicting user interests in online communities.

Knowledge and Information in Society

The various contexts in which the term “commercial content moderators” is used illuminate several problems pertaining to knowledge and information in society. It brings attention to the concealed business-related incentives that influence how user-generated content and other online information is screened and made available to the public. It also illustrates the plight that is faced by knowledge workers in this industry, who are contractually obligated to work long hours for low wages and have minimal or no access to the psychological services that they require. Furthermore, by showing the immense effort and skill that is dedicated to filtering user-generated content, it demonstrates the limitations of technology in this particular field. While algorithms are occasionally used by companies to perform basic forms of moderation, human labour is preferred because people have the ability to make more appropriate judgment calls by discerning the intent of posts based on constantly-evolving cultural norms and values (Roberts, 2016a). Relying on algorithms to do this kind of complex analysis is dangerous, because they cannot interpret intentions in the same way that humans can and would have to be consistently updated to reflect the changing customs and laws in society.

This keyword also makes explicit connections to new ways of understanding knowledge and information in society by revealing the distressing labour that goes into online knowledge curation, which companies try to hide from the public (Roberts, 2016a). By examining the role of commercial content moderators in society, we learn that the information people find online is not objective, and is actually a result of careful deliberation aimed at improving corporate reputation and maximizing profits.

Related Scholars, Policy Events and Key Resources for Further Exploration

Two experts in the area of commercial content moderation are Sarah T. Roberts (2016c), an Assistant Professor from UCLA, and Adrian Chen (2016), a journalist who works for The New Yorker. Roberts (2014; 2016a) wrote her dissertation on this subject, entitled “Behind the Screen: the Hidden Digital Labor of Online Content Moderators,” and recently authored a chapter about content moderation in the book, The Intersectional Internet: Race, Sex, Class, and Culture Online. She has also appeared on radio talk shows such as CBC Ontario Morning and on news programs like TVO’s The
Agenda with Steve Paikin, and features these appearances on her website, *Illusion of Volition* (Roberts, 2016c). During his previous occupation as a journalist at *Wired* magazine, Adrian Chen (2014; 2015) wrote a case study on commercial content moderation in the Philippines titled “The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed.” He has also written an article called “When the Internet’s ‘Moderators’ Are Anything But” for *The New York Times Magazine*.

Policy events that delve into contentious issues within the practice of commercial content moderation include the conference series, “Digital Labour: The Internet as Invisible Sweatshops, Picket Lines, Barricade,” which took place at NYC from November 14-16 in 2014 and can be streamed online, as well as the “digital\human\labour” conference series, which is hosted by the American Association of Geographers and will be held from April 5-9, 2017 in Boston, Massachusetts (The New School, 2014; AAG, n.d.).

**Conclusion**

Overall, the work of commercial content moderators acts as a catalyst for conversation about this multidimensional topic. It brings social awareness to the discriminatory decision-making process used in commercial content moderation procedures by profiteering companies, despite this sometimes being done at the expense of minority groups. It also draws attention to the difficult labour conditions endured by commercial content moderators in developing countries—namely long hours, low wages, and a lack of adequate medical resources for psychological trauma that may be sustained. Further, it illustrates the difficulties that moderators face when balancing human rights to expression and protection. In examining this keyword, I recommend that better regulation be enforced to protect the interests of moderators, minority groups, and users from company exploitation. Moderation procedures should also be modified to preserve potential evidence for legal investigations, as some deleted posts depict criminal activity. Ultimately, the future success and integrity of commercial content moderation lies in the hands of law enforcement officials, policymakers, and the online community which shapes the working environment of moderators and has the power to change the struggles that they face.
References


