The Humans Behind the Filter: Uncovering the Costs and Consequences of Content Moderators in Kenya

Søren Bøgh Sørensen*†[1] and Ephantus Kanyugi*

[1]Copenhagen Business School [Copenhagen] – Denmark

Abstract

Kenya is known as an important destination for the outsourcing of Information Technology
Enabled Services (ITES) like call centres, data generation and annotation, academic writing, product reviewing as well as content moderation. The need for this diverse array of digital labour has only grown in tandem with the proliferation of social media platforms, where graphic, sexually explicit and often disturbing content must be filtered or removed, as well as
the uptake of the development and deployment of machine learning systems, whose outputs must be assessed and verified by human workers. This work is often performed either on online,
digital labour platforms, or in Business Process Outsourcing centers, and Kenya has been the
center of several controversies involving the outsourcing of such tasks by major technology companies like Meta and OpenAI. Due to the initial racist and sexist biases of some of the first
iterations of OpenAI's ChatGPT, e.g., workers were hired through the impact sourcing BPO Samasource to label text passages scraped from the internet to train the Large Language Model
(LLM) to be able to detect toxic material before reaching its users (Perrigo, 2023). Likewise, Meta outsourced content moderation tasks through Samasource, leaving around 140 workers with severe psychological diagnoses like Post-Traumatic Stress Disorder (PTST), Generalized
Anxiety Disorder (GAD), and Major Depressive Disorder (MDD) (Booth, 2024; Gebrekidan, 2024). Given the increasing popularity of the social media platform TikTok in Kenya, where sexually explicit content often appears during nighttime hours, the need for content moderators
to filter through such disturbing content is only increasing (Reuters, 2023; BBC, 2025). For researchers and policymakers, this means that there is an urgent need to understand not only
the psychological consequences but also the broader psycho-social risks, working conditions, and legal issues involved with this type of work. This paper presents some initial findings on these issues in the Kenyan context, highlighting the costs and consequences for the human labour lying behind some of the most popular digital platforms in the global digital economy.

*Speaker
†Corresponding author: sobs.msc@cbs.dk

Keywords: Kenya, Content Moderation, Psycho, social risks