‘It’s destroyed me completely’: Kenyan moderators decry toll of training of AI models

Publication Date: 2023-08-02

Author: Unknown

Section: Technology

Tags: Artificial intelligence (AI), Kenya, Meta, OpenAI, Computing, ChatGPT, features

Article URL: https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai

The images pop up in Mophat Okinyi’s mind when he’s alone, or when he’s about to sleep. Okinyi, a former content moderator for Open AI’s ChatGPT in Nairobi, Kenya, is one of four people in that role who have filed a petition to the Kenyan government calling for an investigation into what they describe as exploitative conditions for contractors reviewing the content that powers artificial intelligence programs. “It has really damaged my mental health,” said Okinyi. The 27-year-old said he would would view up to 700 text passages a day, many depicting graphic sexual violence. He recalls he started avoiding people after having read texts about rapists and found himself projecting paranoid narratives on to people around him. Then last year, his wife told him he was a changed man, and left. She was pregnant at the time. “I lost my family,” he said. The petition filed by the moderators relates to a contract between OpenAI and Sama – a data annotation services company headquartered in California that employs content moderators around the world. While employed by Sama in 2021 and 2022 in Nairobi to review content for OpenAI, the content moderators allege, they suffered psychological trauma, low pay and abrupt dismissal. The 51 moderators in Nairobi working on Sama’s OpenAI account were tasked with reviewing texts, and some images, many depicting graphic scenes of violence, self-harm, murder, rape, necrophilia, child abuse, bestiality and incest, the petitioners say. The moderators say they weren’t adequately warned about the brutality of some of the text and images they would be tasked with reviewing, and were offered no or inadequate psychological support. Workers were paid between $1.46 and $3.74 an hour, according to a Sama spokesperson. When the contract with OpenAI was terminated eight months early, “we felt that we were left without an income, while dealing on the other hand with serious trauma”, said petitioner Richard Mathenge, 37. Immediately after the contract ended, petitioner Alex Kairu, 28, was offered a new role by Sama, labeling images of cars, but his mental health was deteriorating. He wishes someone had followed up to ask: “What are you dealing with? What are you going through?” OpenAI declined to comment for this story. Sama said moderators had access to licensed mental health therapists on a 24/7 basis and received medical benefits to reimburse psychiatrists. In regards to the allegations of abrupt dismissal, the Sama spokesperson said the company gave full notice to employees that it was pulling out of the ChatGPT project, and were given the opportunity to participate in another project. “We are in agreement with those who call for fair and just employment, as it aligns with our mission – that providing meaningful, dignified, living wage work is the best way to permanently lift people out of poverty – and believe that we would already be compliant with any legislation or requirements that may be enacted in this space,” the Sama spokesperson said. The human labor powering AI’s boom Since ChatGPT arrived on the scene at the end of last year, the potential for generative AI to leave whole industries obsolete has petrified professionals. That fear, of automated supply chains and sentient machines, has overshadowed concerns in another arena: the human labor powering AI's boom. Bots like ChatGPT are examples of large language models, a type of AI algorithm that teaches computers to learn by example. To teach Bard, Bing or ChatGPT to recognize prompts that would generate harmful materials, algorithms must be fed examples of hate speech, violence and sexual abuse. The work of feeding the algorithms examples is a growing business, and the data collection and labeling industry is expected to grow to over $14bn by 2030, according to GlobalData, a data analytics and consultancy firm. Much of that labeling work is performed thousands of miles from Silicon Valley, in east Africa, India, the Philippines, and even refugees living in Kenya's Dadaab and Lebanon's Shatila – camps with a large pool of multilingual workers who are willing to do the work for a fraction of the cost, said Srravya Chandhiramowuli, a researcher of data annotation at the University of London. Nairobi in recent years has become a global hotspot for such work. An ongoing economic crisis, matched with Nairobi's high rate of English speakers and mix of international workers from across Africa, make it a hub for cheap, multilingual and educated workers. The economic conditions allowed Sama to recruit young, educated Kenyans, desperate for work, said Mathenge. "This was our first, ideal job," he said. During the week-long training to join the project, the environment was friendly and the content average, the petitioners said. "We didn't suspect anything," said Mathenge. But as the project progressed, text passages grew longer and the content more disturbing, he alleged. The task of data labeling is at best monotonous, and at worst, traumatizing, the petitioners said. While moderating ChatGPT, Okinyi read passages detailing parents raping their children and children having sex with animals. In sample passages read by the Guardian, text that appeared to have been lifted from chat forums, include descriptions of suicide attempts, mass-shooting fantasies and racial slurs. Mathenge's team would end their days on a group call, exchanging stories of the horrors they'd read, he said. "Someone would say your content was more severe or grotesque than mine and so at least I can have that as my remedy," he said. He remembers working in a secluded area of the office due to the nature of the work: "No one could see what we were working on," he thought. Before moderating content for OpenAI's ChatGPT, Kairu loved to DJ. Be it at churches or parties, interacting with different groups of people was his favorite part of the job. But since reviewing content from the internet's darkest corners for more than a six-month period he has become introverted. His physical relationship with his wife has suffered, and he's moved back in with his parents. "It has destroyed me completely," he said. Several of the petitioners said they received little psychological support from Sama, an allegation the company disputes. "I tried to reach out to the [wellness] department to give indication of what exactly was taking place with the team, but they were very non-committal," said Mathenge. Okinyi said counselors on offer didn't understand the unique toll of content moderation, so sessions "were never productive". Companies bear significant responsibility According to its website, "Sama is driving an ethical AI supply chain that meaningfully improves employment and income outcomes." Its clients include Google, Microsoft and Ebay, among other household names, and in 2021 was one of Forbes's "AI 50 Companies to Watch". The company has workers in several places in east Africa, including more than 3,500 Kenyans. Sama was formerly Meta's largest provider of content moderators in Africa, until it announced in January it would be "discontinuing" its work with the giant. The news followed numerous lawsuits filed against both companies for alleged union-busting, unlawful dismissals and multiple violations of the Kenyan constitution. Sama canceled its contract with OpenAI in March 2022, eight months early, "to focus on our core competency of computer vision data annotation solutions", the Sama spokesperson said. The announcement coincided with an investigation by Time, detailing how nearly 200 young Africans in Sama's Nairobi datacenter had been confronted with videos of murders, rapes, suicides and child sexual abuse as part of their work, earning as little as $1.50 an hour while doing so. But now, former ChatGPT moderators are calling for new legislation to regulate how "harmful and dangerous technology work" is outsourced in Kenya, and for existing laws to "include the exposure to harmful content as an occupation hazard", according to the petition. They also want to investigate how the ministry of labor has failed to protect Kenyan youth from outsourcing companies. Kenya's ministry of labor declined to comment on the petition. But companies like OpenAI bear a significant responsibility too, said Cori Crider, director of Foxglove, a non-profit legal NGO that is supporting the case. "Content moderators work for tech companies like OpenAI and Facebook in all but name," Crider said in a statement. "The outsourcing of these workers is a tactic by tech companies to distance themselves from the awful working conditions content moderators endure." Crider said she did not expect the Kenyan government to respond to the petition anytime soon. She wants to see an investigation into the pay, mental health support and working conditions of all content moderation and data labeling offices in Kenya, plus greater protections for what she considers to be an "essential workforce". Beyond the petition, glimpses of potential regulation are growing. In May, the first trade union for content moderators in Africa was formed, when 150 social media content moderators from TikTok, YouTube, Facebook and ChatGPT met in Nairobi. And while outsourced workers are not legal employees of their clients, in a landmark ruling last month, employment court judge Byram Ongaya ruled that Meta is the "true employer" of its moderators in Kenya. It remains unclear to whom OpenAI currently outsources their content moderation work. To move forward, it helps Okinyi to think of ChatGPT's users that he has protected. "I consider myself a soldier and soldiers take bullets for the good of the people," he says. Despite the potential for bullet wounds to stay forever, he considers himself a hero.