Preface

These are the stories of chat moderators and data annotation workers. My name is Michael Geoffrey Abuyabo Asia. I am an ex-chat moderator and data worker who has served across multiple global outsourcing platforms. I worked on Meta-related projects under Sama and later took on roles at CloudFactory, TELUS International, TransPerfect DataForce, Appen, and NMS Philippines. In much of this work, I was tasked with impersonating and training AI companions—performing emotional labor that required intense psychological presence while remaining entirely invisible to the users who interacted with these systems.

Like many of us, I entered this work through subcontracting structures that deliberately obscure responsibility. The labor is fragmented, undervalued, and consistently portrayed as "simple," even as it exposes workers to disturbing content and demands finely tuned emotional performance. These narratives of "low-skill" work hide a deeper reality: a rapidly expanding industry that relies on hidden human labor while offering minimal protections, unstable contracts, and little recognition.

This report is an account of my own experiences—experiences widely shared by others doing this work. For this study, I interviewed seven of my colleagues, drawing on their testimonies and my lived experience to document the pressures, harms, and unresolved grievances shaping this sector.

Funded by the Distributed AI Research Institute (DAIR), the Weizenbaum Institute, and Technische Universität Berlin, the Data Workers' Inquiry foregrounds the voices and demands of data workers. It seeks to prevent further violations, establish the facts of this evolving industry, and foster the solidarity needed for meaningful change.

Acknowledgements

TRIGGER WARNING

This report includes depictions of stressful working conditions, mental health burdens, and structural injustices that may be distressing for some readers.

Viewer discretion is advised.
 Please proceed with care.

The path to chat moderation

After graduating as an air cargo agent from Nairobi Aviation College in Kenya, I dreamed of charting global skies; mapping intricate routes for cargo flights, decoding the chaos of aviation logistics. But fate had other flight plans. Finding a job in the aviation sector proved nearly impossible. There were few opportunities, fierce competition, and endless rejection. Bills piled up, my confidence sharply declined, and the silence from potential employers became deafening. I had invested my precious time, energy, and passion into completing the right courses, networking, and applying to countless positions, to no avail. I felt invisible. I tried to stay optimistic, reapplying and eagerly awaiting job interviews, but doors just did not open. Gradually, what had started as a purposeful journey felt like wandering in endless circles. Desperation crept in. I felt like I couldn't wait for that perfect role in aviation. I had to survive.

As a father with young children, I set aside my career ambitions and focused on providing for my family. That's how I ended up working for Samasource (now known as Sama). I was stationed at their office in Lavington, Nairobi. There, I was taught how to label text, images, audio, and video to help train AI systems. I was unaware that this new digital path carried its own hidden costs, including crazy key performance indicators (KPIs) which included, but were not limited to, billable hours, average handling time, occupancy, and quality assurance, amongst others. The poor pay, which left me unable to cater to my fundamental needs pushed me to find alternative sources of income.

It was during this period, just after I had been absorbed as a data labeler at Sama, and after countless conversations with colleagues, that I was introduced to the world of chat moderation, something I didn't know existed but was soon to change my life. Chat moderation refers to a situation where you have interactions with users, in my case, romantic and intimate conversations on platforms I'd never heard of. Chat moderation jobs were advertised on platforms such as TextingFactory, e-moderators, Cloudworkers, remotely4u, and the one I ended up working for, New Media Services.

Below is a typical screenshot for a text chat operator job on New Media Services.
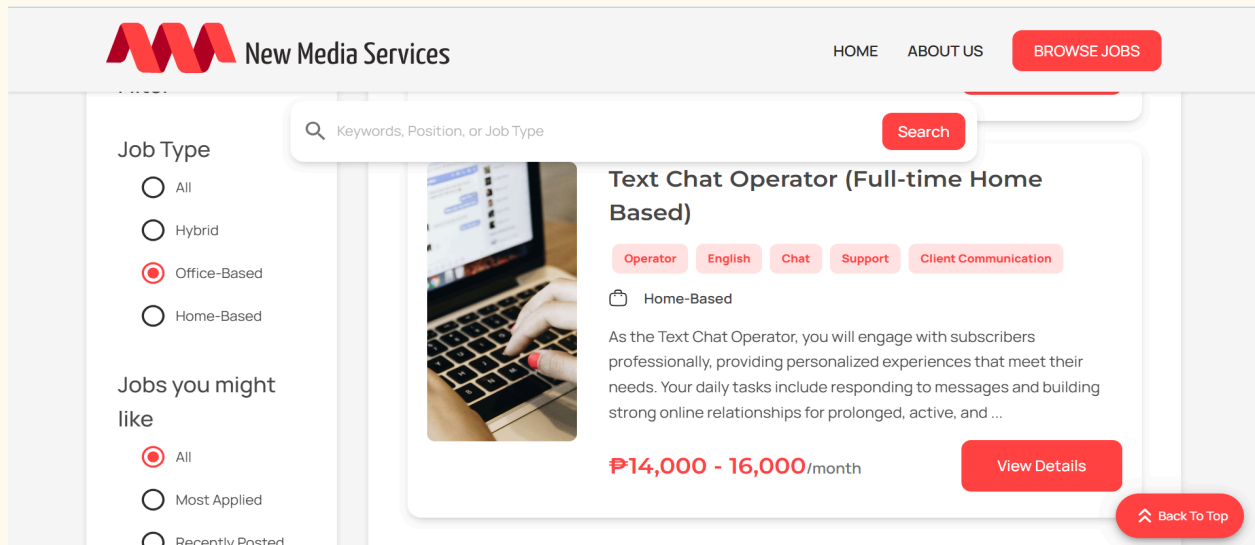
The role was not clearly defined. The job description said a text chat operator was supposed to facilitate what they referred to as "interactive and creative communication" with customers and subscribers through instant messaging, and to provide a "personalized experience that meets their needs." What I didn't know was that the role would require me to assume multiple fabricated identities, and use pseudo profiles created by the company to engage in intimate and explicit conversations with lonely men and women. The job title "chat moderator" gave no hint of what it truly entailed.

I had to assume fabricated identities, memorizing false backstories, and reading through previous chats. Sometimes I would be assigned a conversation that had been ongoing for several days and had to continue it smoothly so the user wouldn't realize the person responding had changed. I played the part, stepping into carefully crafted personas designed to connect with unsuspecting customers on a "personal" level, often through sexual or intimate conversations.

When I logged into my work dashboard, I had access to multiple fake profiles of varying genders —typically three to five different personas I could operate simultaneously. Sometimes I had to operate male and female personas on the same day, depending on what the platform's users were seeking. All profiles would have an avatar, and some would have normal names like Peter, Susan, James, etc, while some would have

usernames like Sexydoll124. Sometimes, nobody would inform me of the race or gender of the profile, and I would need to scan the previous chats to figure it out.

One day, I might be Jessica, a 24-year-old lesbian college student from California, and Joe, a 30-year-old gay man from Florida. Another day, it could be Maria, a 28-year-old heterosexual nurse, or a nameless woman artist. The variety was designed to cater to diverse user preferences — some wanted to chat with women, others with men — and the platform needed moderators who could convincingly portray both.

During peak hours, I juggled conversations across these different personas. Before responding, I had to quickly review each chat history to recall not only the personality I was portraying but the details I'd shared and the storyline I was stepping into. Because several moderators could operate the same profile across different shifts, I often had to pick up from conversations started by someone I'd never met and ensure the user never noticed the handover. Switching from being a flirty female student in one window to a confident male trainer in another required entirely different writing styles and emotional tones. The challenge was maintaining authentic-seeming conversations regardless of who users believed they were talking to.

The company required us to ask personal questions, remember details about users' lives, and respond in ways that made them feel heard and desired. I spent hours crafting messages meant to make strangers feel special, wanted, even loved. They confided in me about marriages, divorces, and fears of dying alone, and I replied with scripted empathy, manufactured attraction, and choreographed intimacy.

What the job description didn't tell me was how deeply this work would conflict with everything I believed about honesty and dignity. My faith taught me that love should be real, intimacy sacred, and that deception was destructive to both the liar and the deceived. Yet here I was, professionally deceiving vulnerable people who were genuinely looking for connection—taking their money, their trust, their hope, and giving them nothing real in return. But I was already in, and bills don't wait for convictions.

The reality behind the job

In my tiny single room in Mathare slums, Nairobi — a place I couldn't even afford at the time, with rent already weeks overdue — I found myself living a double life. As a man, I

had to provide shelter, clothing, food, and many other basic needs to my young family. No excuses. No mercy. It was my obligation. Right beside me, my wife and kids were sound asleep, believing I was fixing broken servers, responding to harmless tech tickets. Little did they know that I had just told another man, "I love you."


Mathare slums at a glance

The company's non-disclosure agreement (NDA) ensured I couldn't tell them the truth. It was mandatory. You had to consent to the NDA to continue with the application process. What initially seemed like a standard privacy protection became a tool of silence and control. I couldn't disclose the nature of the work, the fake identities, the sexual content, or the psychological toll. The NDA didn't just protect the company's business model; it isolated us, preventing us from seeking support or even explaining to our families what we were actually doing for money.

The platform's rules were absolute: be convincing, be available, but never be real. We were strictly prohibited from sharing any actual personal details with users, no real names, no real locations, no real phone numbers, nothing true about ourselves. Every interaction had to remain within the fabricated persona. I became an expert at telling people exactly what they needed to hear while my own identity was completely erased. I was supposed to be their confidant, their fantasy, and their emotional support, even as I was crumbling inside.

My sole responsibility was to keep the conversation going, no matter what it did to me. The users wanted a deep connection, the company wanted their numbers up, and I was bleeding out inside with every fake message I sent. I was being torn apart in three directions: faking my identity, hiding what I was really doing from my wife without breaking the NDA, and somehow not losing my marriage in the process. If I told my wife that sexting and flirting with strangers was literally my job, not some mistake, but what they actually paid me to do, I was sure she would look at me like I was insane. No one would believe it. Not just her. How do you explain that you get paid to tell strangers you love them while your real family sleeps three meters away?

https://www.awesomescreenshot.com/video/41968549?key=1422ce32399e372053166b38fc02bc57.
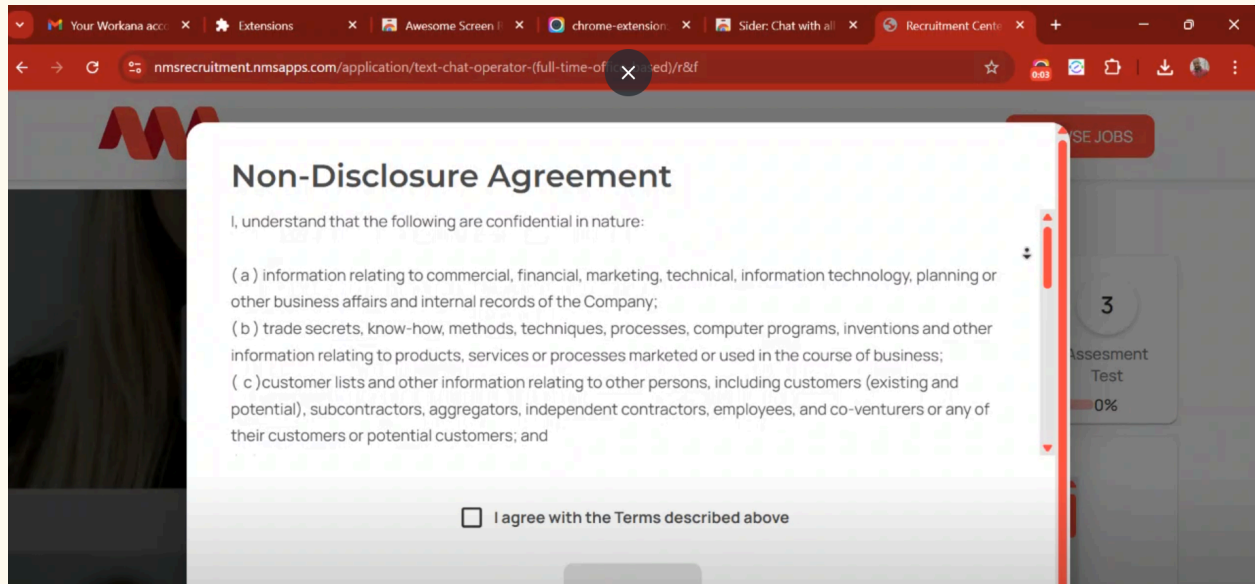
The challenge went beyond just switching genders; I also had to perform across different sexual identities. Sometimes I was a straight woman flirting with men. Other times, I had to be a lesbian woman talking to other women about desires I'd never felt. In some conversations, I was expected to be a gay man discussing intimate details I knew nothing about. As a straight man, I had no roadmap for understanding what these users wanted or needed, but I had to fake it anyway. Bills don't wait for your comfort zone. The guilt ate at me every night.

I felt like I was losing myself in the role. It started as any other job, responding with empathy and willfully pretending to care, but over time, it became harder to separate the act from reality. The lines blurred. I began questioning if I was acting or if I was truly becoming the persona I was forced to embody. Every moment of pretense fractured something inside my spirit, and my sense of self. I was losing touch with who I really was, a feeling that has never left me. Working such a role in the morning and then later running church affairs as a service programmer clashed. I served on the altar, innocently discharging my mandate, but deep down dying of guilt.

Exploiting intimacy for profit

Was the exploitation of intimacy for profit real? Yes. We weren't just paid for messages, but for the illusion, twisting our emotions into revenue. The entire business model depended on creating deeper, more intimate connections with users. The more engaging and creative a chat moderator was, the more the company earned from each conversation. Payment was tied to output: $0.05 per message, each meeting a required character count. To keep up, we had to type at least 40 words per minute, a skill tested during the recruitment process, while juggling multiple chats simultaneously. The platform dashboard displayed live tallies of messages sent, a constant reminder that every line of manufactured affection, every fake confession, every scripted "I love you" had a price tag.

The company also provided pornographic stickers, which we suspected were generated through their partner websites, though these affiliations were never clearly disclosed and were hidden behind strict non-disclosure agreements. Below is a screenshot showing how applicants were required to consent to interacting with adult content on Texting Factory, one of the platforms I worked on. Could we have been flirting with users from platforms like Ashley Madison, BeNaughty, Flirt.com, Tinder, MenNation, or Grindr? The company never confirmed it, but given the overlapping content and tone, it remains a strong possibility.

I quickly realized that this was not ordinary chat work but a tightly monitored production system. Every interaction was tracked through key performance indicators (KPIs), including message volume, response time, user engagement, and continuity. Meeting these targets was mandatory.

We were expected to sustain long, personalized conversations across multiple users and shifts without breaking character. Our accuracy, creativity, and typing speed were constantly measured. Falling behind on metrics could lead to warnings, reduced assignments, or termination. What seemed like a simple chat role was, in reality, a high-pressure system of continuous performance; monitored, quantified, and monetized in real time.

Was I training AI companions?

I always suspected that some of the people on the other side of the chats thought I was an AI companionThey'd sometimes ask if I was real, or test me with questions that seemed designed to catch a bot. When I later read about AI companions, it hit me: the company was probably using me to train these systems. Every conversation I had, every response I crafted, every way I learned to keep users engaged, all of that data was likely feeding into algorithms that would eventually replace me. I was training my own replacement, teaching machines how to manipulate lonely people the same way I was being forced to. The moment I logged in under that fake name, the platform completely

erased my human limits and morals. The user wasn't talking to the real me at this point, yet I was the one drowning in their filth.

But the confusion ran deeper than that. I began to wonder: what if I wasn't just training an AI companion, what if I was actually impersonating one? Maybe users thought they had already purchased an AI girlfriend or boyfriend, and I was the human pretending to be the machine pretending to be human. That would explain why the personas were so stripped of real detail, why everything had to remain surface-level and fantasy-driven. The company never clarified, and perhaps that ambiguity was intentional. It didn't matter to them whether users believed they were talking to AI or a real person, as long as they kept paying.

Looking back, it makes sense. Every message I sent was recorded. The platform tracked how fast I replied, which words kept users engaged, and what tone worked best. It felt like the company was collecting more than just labor, they were collecting patterns: how we joked, comforted, or flirted. All that data could easily be reused to build chatbots that sound more human.

I also began to suspect there was a middle platform between me and the end user. It never made sense how I could be suddenly assigned to an ongoing conversation and step into a profile so seamlessly. What was the user actually seeing on their side? Were they interacting directly with me, or through another interface that masked the handovers? Every chat moderator was identified by a message code, a small tag that marked which responses came from you. That code helped me trace where my messages started and where another moderator's ended. But if users couldn't see those codes, it meant there was an entire layer of the system mediating our exchanges, deciding which operator picked up which chat, and keeping the illusion of one continuous "person" alive. It made me wonder if the system itself was learning from these transitions, using us to test how convincingly human a conversation could remain across multiple hands.

The psychological toll of this confusion was devastating. Then came the moment that showed me how far gone I was. The platform would periodically test us to ensure we weren't bots; they'd ask us to prove we were real humans behind the screen. But I had been living inside these fabricated personas for so long that when they asked me to prove my humanity, I panicked and started defending whichever fake identity I was currently operating instead. I caught myself explaining why I was a "real woman," talking about "my" periods, "my" dating experiences, "my" feelings about men, instead of just proving I was a human being. I was so deep in the character that when challenged, I defended the

lie instead of the truth. The fake personas had become so automatic that I forgot the test wasn't about gender; it was about whether I was human or AI. I was more committed to defending a lie than I was to being myself.

Consent and boundaries

Here is an example of a message I received from a user. On paper, it may not seem offensive, but for me, it carried a different kind of discomfort. As a straight man, being asked by another man for sex and having to respond as if I were a willing woman created a deep sense of revulsion and helplessness. I wanted to stop, to close the chat, to walk away—but I couldn't. The platform's rules made it impossible to refuse or skip conversations. My job was to remain polite, seductive, and available, regardless of how uneasy I felt.
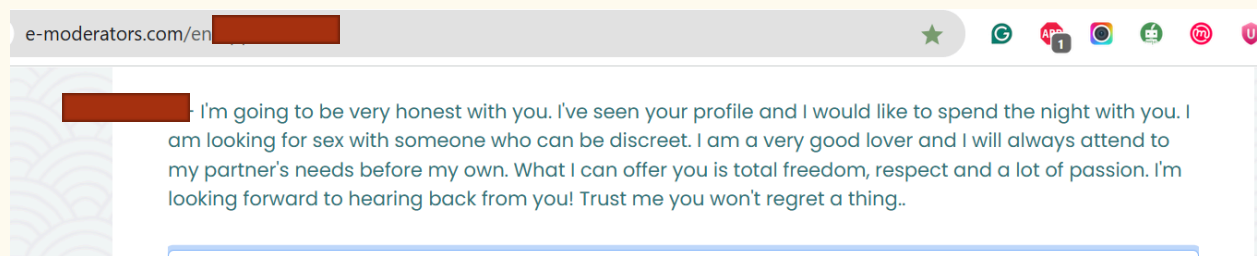
The platform stripped me of my right to say no. I became a complete slave to the system, complying with every command. The user wasn't talking to the real me anymore; I was just a proxy, performing someone else's fantasy. The work itself felt filthy and suffocating, yet the thought of being replaced by a machine that could do it "better" wasn't comforting either. I wasn't just exploited, I was also disposable.

Each instance of compelled affection created invisible but internal wounds. I was reduced to a voice without a will, a presence without consent. Every night, after the shift came to an end, I couldn't even look my wife in the eyes without it leaving a mark on me. I could barely sleep knowing I had lied. I had to sideline my religious beliefs and principles throughout the shifts, telling lies for hours. Every time I stepped into church,

my eyes would always land on the same scripture — it felt like it was staring straight into my soul: *"Jesus is the way, the truth, and the life."* John 14:6

The psychological damage wasn't contained to my work shifts. My wife started noticing the difference in my character. I began isolating myself, trying to push away those I loved. I could not withstand the guilt of telling my wife I loved her when I had really lost the taste of love and had no love deposits left. There were nights I could barely recognize the stranger behind the laptop. In the morning, I would speak to my beloved wife and wonder if I still had any truth left in me, or if every word that came out of my mouth was just another fabricated script.

A fellow chat moderator I interviewed for this essay told me, "It's painful how our morals and values are completely damaged in the name of work." I could feel this colleague's pain from her facial expressions. "My entire service to ministry work was by far affected by this 'hell of a job,'" she continued.

"Over time, the damage has remained very evident, even from my social life," another chat moderator I spoke with told me. "Love is fake. If there is any scope of love left, maybe it's in the romantic movies." Threats of a breakup with my partner eventually became a reality. The disconnection turned disastrous,she left, together with my two-year-old son."

Human rights over algorithms: demands for ethical AI

Over the years, the emotional toll has been rampant and gone unaddressed. But this doesn't mean no one was watching. We were watching. We felt every wound. And now we're speaking up.

What happened to me, and to thousands of others performing this invisible labor, was not just exploitation. It was a systematic erasure of our humanity in service of building the very technologies that would replace us. We were used to teaching machines how to mimic intimacy, how to exploit loneliness, how to sound human while remaining utterly hollow. And we were silenced through NDAs, isolated by shame, and discarded when we could no longer perform.

If AI systems are going to be built on the backs of human workers, then those systems must be held accountable to human welfare. I demand transparency about how these platforms operate: full disclosure of AI architectures, the hidden connections between chat services, AI companions, and dating platforms, and the monetization frameworks

that profit from manufacturing fake intimacy. It is no longer defensible to secretly manipulate human behavior.

I demand that AI systems be evaluated for their psychological, emotional, and societal consequences before deployment, not after the damage has been done. Profit-driven models must prioritize human emotional, cognitive, and social welfare above revenue. There must be independent, empowered ethical review boards with the authority to assess, monitor, and halt practices that undermine human welfare or exploit psychological vulnerabilities.

And I demand that workers, the actual human beings performing this labor, have a voice in governance. We are not disposable. Our experiences matter. Our testimony matters. The community of workers and the people affected by these systems must be incorporated into decisions about how AI is built, developed, deployed, and regulated.

This is not just about better working conditions. This is about recognizing that you cannot build ethical AI on the foundation of exploited, traumatized human labor. If the training data comes from people who were psychologically destroyed in the process of creating it, then the technology itself is tainted from the start.

We deserve better. The users being deceived deserve better. And the future of AI must be built on something other than our broken humanity. Until then, remember that an AI girlfriend responding to your loneliness might just be a man in a Nairobi slum, wondering if he'll ever feel real love again.

Data Workers' Inquiry