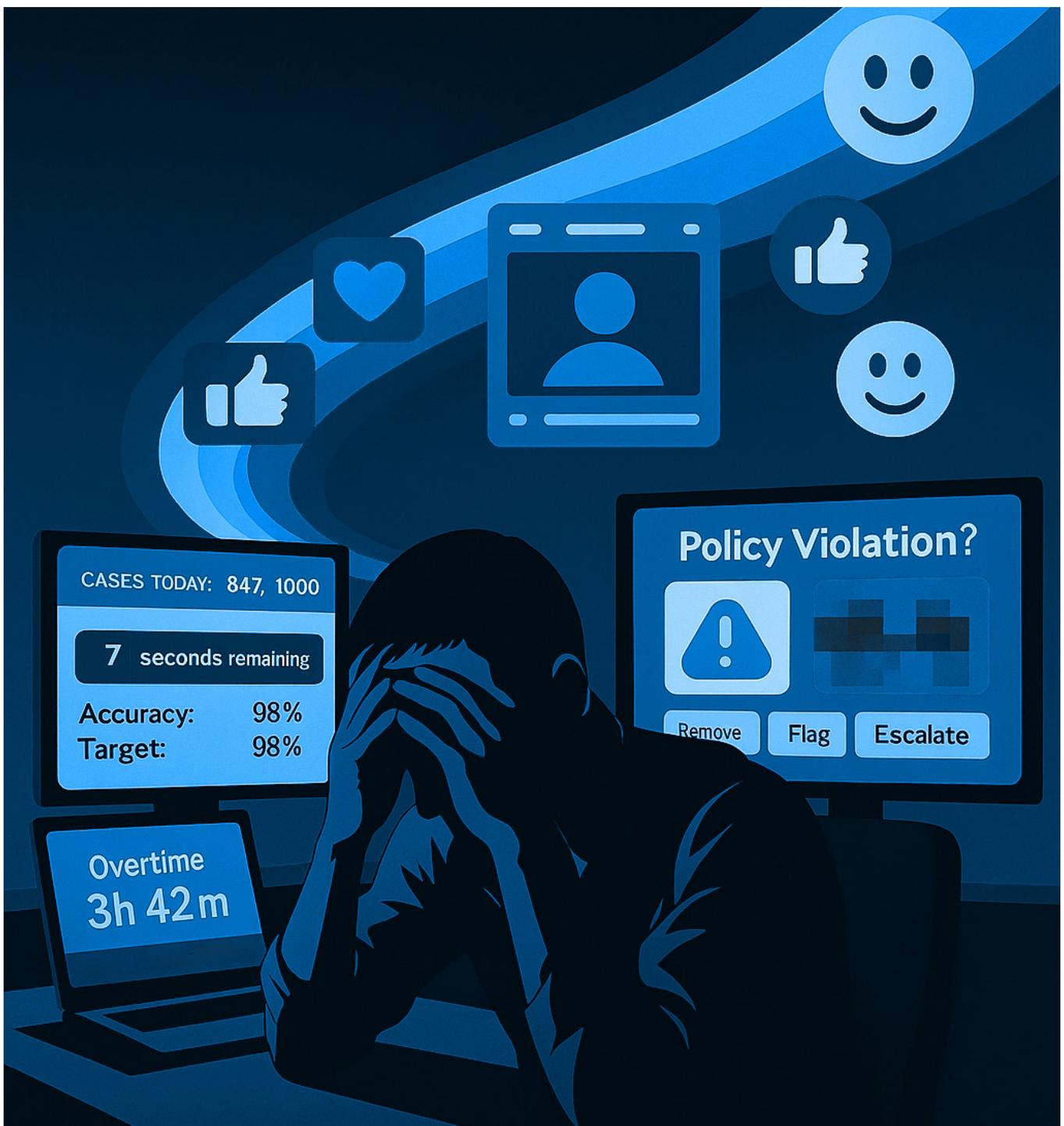# SCROLL. CLICK. SUFFER.

The Hidden Human Cost of Content Moderation and Data Labelling

equidem

EQUIDEM

Equidem is a human rights organization that is rightsholder-centred and representative of the Global Majority. We work both globally and locally to expose injustice and hidden human rights abuses; to provide practical, evidence-based solutions to some of the world's most intractable human rights challenges; and to build powerful, community-led coalitions that empower individuals and communities to claim their rights.

equidem.org

CONTENT WARNING

This report contains first-hand accounts of violence and harassment, including instances of sexual abuse, experienced by data labellers and content moderators. Reader discretion is strongly advised. We recommend that survivors and others who may be affected by this content take the necessary steps for emotional preparedness before reading, and seek support if needed while engaging with this material.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

3

TABLE OF CONTENTS

4   Scroll. Click. Suffer
         The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

5

ABBREVIATIONS AND ACRONYMS

BPO	Business Process Outsourcing
CSDD	Corporate Sustainability Due Diligence
EU	European Union
EU CSDDD	EU Corporate Sustainability Due Diligence Directive
GBV	Gender-based Violence
GBVH	Gender-based Violence and Harassment
HR	Human Resources
HRDD	Human Rights Due Diligence
IIC	Inappropriate Interactions with Children
ILO	International Labour Organization
MNE	Multinational Enterprises
NDA	Non-disclosure Agreements
PTSD	Post-traumatic Stress Disorder
OECD	Organisation for Economic Co-operation and Development
OSH	Occupational Safety and Health
UNGPs	UN Guiding Principles on Business and Human Rights

TABLES AND IMAGES

FIGURE 1: Overview of business and employment relationships between social media platforms and their supply chains

TABLE 1: Select details about multinational technology companies that have operation in EU markets

TABLE 2: Occupational safety and health impacts of ongoing exposure to violent and explicit content

TABLE 3: Economic harms faced by data workers

TABLE 4: Incidences of harm reported by workers, listed by country

TABLE 5: Informal hiring practices that exclude data workers from labour rights protections

TABLE 6: Violations of ILO and human rights due diligence standards

TABLE 7: Distribution of workers engaged by nationality and role

TABLE 8: Informal hiring practices documented by Equidem on digital technology supply chains in Colombia, Kenya, and The Philippines

TABLE 9: Physical harms reported by data workers

TABLE 10: Incidences of physical harms reported by data workers

TABLE 11: Psychological harms reported by data workers

TABLE 12: Incidences of psychological harms reported by data workers

TABLE 13: Sexual harms reported by data workers

TABLE 14: Incidences of sexual harms reported by data workers

TABLE 15: Economic harms reported by data workers

TABLE 16: Incidences of economic harms reported by data workers

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

7

EXECUTIVE SUMMARY

> *My friend was punished because she was out of her working station. She was in such a bad mental state that she did not fight back. She had a mental breakdown and left crying. She did not take a break or anything – she just had to leave, and she did. She took two hours crying with the psychologist, and when she returned, she had an action form waiting for her. After that, she just quit.*

Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance[1]

Scrolling through user-generated content on social media is a routine experience for 5.2 billion people – or 63.9% of the global population.[2] Most users take for granted the clean, sanitised, and curated feel of their feeds, rarely pausing to consider what's been removed. But behind these seamless experiences lie an invisible workforce: data workers who manually sift through, label, and flag horrific violence and graphic sexual exploitation from an endless stream of disturbing content. The rampant proliferation of user-generated content on platforms like Facebook, Instagram, Threads, and TikTok has brought with it a flood of explicit imagery depicting violence and exploitation – material that is a risk to public health and safety. Content moderators provide this essential public service, together with data labellers who assign labels to raw data to help provide context for machine learning, including for popular products like OpenAI's ChatGPT. As explained by Dr. Sarah Roberts, Professor and Director of the Center for Critical Internet Inquiry at UCLA, people are involved in every step of developing and implementing AI models: *[T]here are humans in the loop at every step of the way. Somebody had to program that algorithm, somebody tested it. People train data for it by labelling data sets. So, the idea that we might be in a moment at some point in the near future where humans are not involved in this activity is totally aspirational.*[3]

This report presents the findings of Equidem's investigation into the consequences of ongoing exposure to the world's most obscene images of violence and exploitation for data enrichment workers ("data workers"), including content moderators and data labellers. Our researchers engaged with 116 data workers distributed across Colombia, Kenya, and The Philippines, including workers employed by multinational digital solutions firms and third-party BPO companies for multinational technology companies, including Meta, ByteDance, and OpenAI. We found that rather than receiving heightened protection as essential workers, data workers on content moderation and data labelling value chains conduct dangerous work – without protection and under exploitative conditions.

Content moderators:
Content moderators review user-generated content on online platforms to ensure it complies with the platform's guidelines and rules. They identify and remove or flag content that is offensive, harmful, or violates community standards, essentially helping to maintain a safe and positive online environment.

Data labellers:
Data labellers add labels or annotations to data, making it usable for training machine learning models. They essentially provide the "answers" or context that a model needs to learn and make predictions, including recognising violent and sexually explicit content. This process is crucial for tasks like object recognition in images, understanding spoken words in audio, or identifying patterns in text.

8   **Scroll. Click. Suffer**
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org
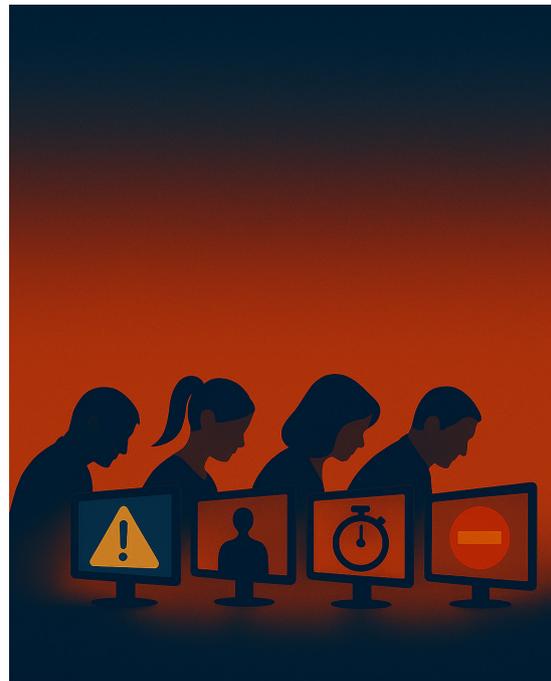
Table 1: Select details about multinational technology companies that operate in EU markets

	Selected Products	Headquarters	Operation in EU Markets
Meta	Facebook Instagram WhatsApp Threads	United States	Yes
ByteDance	TikTok	China	Yes
OpenAI	ChatGPT	United States	Yes

Equidem documented extreme occupational safety and health risks, amounting to workplace violence – including industry conditions that inflict physical, psychological, and sexual harms. Data workers are made to endure constant exposure to graphic, traumatic content with minimal health protections, conditions that violate International Labour Organization (ILO) and human rights due diligence standards under the German Supply Chain Act, OECD Guidelines, emerging EU Corporate Sustainability Due Diligence Directive (CSDDD), and the UN Guiding Principles on Business and Human Rights (UNGPs).

The demand for data workers is on the rise, and the extreme consequences to their health are systemic. The impact of these labour processes is not limited to individuals, but has ripple effects on families, communities, and healthcare systems. For countries increasingly reliant on these streams of work, failure to meaningfully address the risks content moderators and data labellers face is a public health catastrophe in the making.

Table 2: Occupational health and safety impacts of ongoing exposure to violent and explicit content

PHYSICAL HARM	PSYCHOLOGICAL HARM	SEXUAL HARM
Exhaustion	Aggression	Employer refusal to address workplace sexual harassment
Eye strain	Anxiety	Impact on intimate relationships
Hair loss	Decision fatigue	Sexual harassment and violence in the workplace
Insomnia	Depression	Sexual withdrawal
Nightmares	Detachment and indifference	Trauma from labelling and moderating sexually explicit content
Pain in the feet, back, and wrists	Emotional fatigue	
Psoriasis[4]	Irritability	
Substance dependence	Nightmares	
Weight loss	Night terrors	
	Numbness	
	Panic attacks	
	Post-traumatic stress disorder	
	Social withdrawal and isolation	
	Strain on family relationships	
	Stress from targets and penalties	
	Substance dependence	
	Suicidal ideation	
	Traumatic memory recall	

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

9

MODERATING VIOLENT CONTENT

> *I work during the graveyard shift. I come to the office the evening before my shift starts and bring my stuff to sleep in the office because there's no more transportation at dawn when my shift finishes. I receive videos in a queue. Usually the videos contain violence, scams, or are made with underage users. The worst one that I remember was someone being killed. He was tortured on video until he died.*
>
> -
> Pamela, a Filipina national working as a content moderator for a BPO firm subcontracted by a social media platform in The Philippines[5]

The labour of screening endless streams of violent, sexually explicit, and traumatic content isn't just exposure to violence – it is a form of violence in itself. This work inflicts real, lasting harm on the bodies and minds of data workers. It destabilises emotional health, damages psychological well-being, and leads to long-term conditions like PTSD, depression, and anxiety. This is institutionalised harm – a system that normalises violence as part of the workflow and trauma as part of the job. Camilia, working as a content analyst for a BPO firm in Colombia subcontracted by Meta and ByteDance, explained: ***It really depends on what disturbs you. For me, it is child abuse and animal abuse – we can see that – cadavers (and their violation), and human trafficking, sexual solicitation, and prostitution. All of this can be disturbing.[6]***

For data workers, exposure to explicit content is relentless. Workers employed as content moderators and data labellers by BPO firms in The Philippines for social media platforms told Equidem that they were required to work up to 18 or even 20 hours a day. Some employees told Equidem researchers that they had mere seconds to react. Mile, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, described having just seconds to react to a queue of 1,000 images a day: ***We have to moderate a lot of cases. In a normal shift, we have to go through 700 to 1,000 cases, which means we have to accomplish an average time of 7-12 seconds to moderate each case.[7]***

As described in the sections that follow, the very nature of the work of moderating and labelling explicit content poses occupational health and safety risks to data workers – with extreme physical and mental impacts inextricable from the required labour processes.[8] This is not just a labour issue – it's a health crisis.

OVERWORK

> *This environment is becoming more stressful and toxic than even the actual content I'm moderating.*
>
> -
> Evans, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance[9]

The violence of content moderation and data labelling labour processes is compounded by employment insecurity and industrial discipline practices used to drive accelerated work. Workers sift through and sanitise violent content while under threat that if they do not meet targets, they will face pay cuts or forfeit target-based bonuses that can comprise up to 70% of their wages. Data workers told Equidem researchers that they were forced to work unpaid overtime and faced workplace bullying, especially in cases where they had team performance targets.

Aziz, a Kenyan national working as a quality analyst for a BPO firm in Kenya subcontracted by Meta and OpenAI, explained: ***The company pressures us to meet the weekly target. We have to work unpaid overtime to meet them and if we don't meet a target, the company cuts our pay.[10]***

Moderating violent and explicit content at an accelerated pace under the threat of penalty inflicts physical, psychological, sexual, and economic harm on data labellers and content workers – occupational health and safety consequences that amount to violence under ILO Violence and Harassment Convention, 2019 (No. 190).

PHYSICAL HARMS TO DATA WORKERS

Workers on digital technology supply chains told Equidem investigators about the physical consequences of their sustained exposure to explicit violence and sexual content – including depression, exhaustion, hair loss, insomnia, loss of appetite, nightmares, psoriasis, substance dependence and weight loss. These physical harms are heightened in a culture of overwork where data workers are penalised for taking leave, including sick leave.

Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, described the physical health consequences of his work moderating chats for Facebook and Instagram: ***The content was 99% child pornography and abuse material. I was dealing with inappropriate content day in and day out, 8 hours a day. I lost around 15 kilos in 6 to 8 months. I have psoriasis, which has gotten***

10   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

much worse due to stress. I have had terrible sleeping patterns while working there, especially when I was working the graveyard shift. I felt depressed and would drift off, thinking of videos I had seen.[11]

PSYCHOLOGICAL HARM

The exposure to violent and sexual content without adequate mental health support, as well as target-based pressures that drive overwork, verbal abuse, and other forms of workplace bullying, inflicts devastating psychological harms on data labellers and content moderators. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, described the extreme toll of extended hours in this work: *They take the "24-hour availability" way too seriously. It is like you sold yourself to them. They change your shifts, hours, and everything whenever they want to. They don't want you to react, they want you to be so disassociated that you don't stress anymore. And it is good money, so I understand why people stay. I wouldn't be surprised if anyone ever jumped from [the] stairs. We hear people doing drugs in the bathroom any time of the day. It is a horrendous place to be in. I never thought mental health could get this serious.*[12]

SEXUAL HARM

Data workers also confided in Equidem investigators about sexual trauma associated with sustained exposure to pornographic and violent sexual content. Alamini, working for a BPO firm in Kenya subcontracted by Meta and OpenAI, described the impact of prolonged regular exposure on her ability to engage in healthy sexual relationships: *I am constantly exposed to sexually violent content. This has affected my sexual life. I feel detached and no longer have the desire to engage in sexual activity.*[13]

Past research has established that most violent, sexually explicit video content depicts violence against women.[14] Clinical studies on the impact of violent sexual content on women found that women exposed to this content experienced increased anxiety, depression, and PTSD[15], as well as decreased self-esteem and body satisfaction.[16] This disproportionate impact on women data workers constitutes workplace gender-based violence (GBV) under ILO Convention 190.

ECONOMIC HARM

Data workers are also exposed to a slate of practices that inflict economic harm (Table 3).

Table 3: Economic harms faced by data workers

ECONOMIC HARMS		
UNSTABLE EMPLOYMENT	LOW WAGES	UNFAIR WAGE DEDUCTIONS
NO FIXED SALARY	LOW TASK-BASED PAYMENTS	FINANCIAL PENALTIES FOR TAKING SICK LEAVE
FORCED UNPAID OVERTIME	FINANCIAL PENALTIES FOR NOT MEETING TARGETS	FORCED PERIODS OF UNEMPLOYMENT WITHOUT PAY
MANDATORY UNPAID TRAINING		UNFAIR DENIAL OF BONUS PAYMENTS

Iman, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, explained the impact of income inconsistency on her ability to manage family expenses, and the stress associated with these challenges: *Financial stress due to issues with income consistency and unpaid leave have created additional pressure on my family. It has affected our ability to manage expenses and plan for the future.*[17] Economic harms of working in data labelling and content moderation interact with industrial discipline practices that inflict physical, psychological, and sexual harm to heighten experiences of violence for data workers.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

11

WORKPLACE VIOLENCE

The physical, psychological, sexual, and economic harms inflicted on data workers amounts to workplace violence under C190 – Violence and Harassment Convention, 2019 (No. 190). Understanding the experiences of digital supply chain workers as an industry-wide spectrum of violence captures the systematic complex of harms faced by data workers. It also provides a framework for understanding how labour processes and industry practices drive these harms.

Table 4: Incidences of harm reported by workers, listed by country

	Colombia	Ghana	Kenya	Incidents
Physical harms	13	2	16	31
Sexual harms	3	-	7	10
Psychological harms	25	4	31	60
Economic harms	24	6	22	52
Incidents by country	65	12	76	153

This table is based on a qualitative review of 76 worker testimonies from Colombia, Ghana, and Kenya. Each testimony was analysed to identify references to four categories of harm: physical, sexual, psychological, and economic harms. Multiple harms described within a single testimony were individually counted. Where workers mentioned several impacts under the same category (e.g., both anxiety and depression), these were recorded as separate instances. Accordingly, this data represents the total number of harms reported, which notably exceeds the number of unique respondents. This approach aims to highlight not just the presence of harm, but the intensity and prevalence of harms experienced and reported by individual workers. Data from The Philippines was excluded from this table due to a smaller sample of individual testimonies.

FAILURE TO PROVIDE ADEQUATE PSYCHOLOGICAL SUPPORT

The mental health toll on data workers is systemic and severe. While some companies presented wellness programs on paper, in practice these were poorly communicated to workers, ineffective, and inconsistently accessible. In some cases, workers told Equidem they were punished for taking the time to use even the limited mental health resources available. Miguel, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance explained: *[The BPO firm] told us that using more than 1.5 hours of Wellbeing Pauses per week could result in action forms.*[18]

While supportive supervisors can make all the difference to worker wellbeing, supervisors face retaliation for advocating on behalf of workers at the expense of lead firm production targets. Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, told Equidem that as a manager he tried to advocate for his colleagues but was fired and blacklisted for speaking out: *I used my manager's office to advocate for my colleagues, but it backfired on me. It made me get blacklisted and that's what led to my contract not being renewed.*[19]

These failings in providing mental health support are driven by the extreme pressure to meet performance targets at the expense of worker wellbeing. Inability to access meaningful mental health support at work is exacerbated by private health insurance company exclusions from mental health support. For instance, while data workers in The Philippines overwhelmingly reported that health insurance is among the benefits provided for BPO workers, including content moderators and data labellers, most private health insurance companies do not cover mental health treatment.

INFORMALITY AND EXCLUSION FROM LABOUR RIGHTS PROTECTIONS

> *The company called me to threaten me. They talked about how easy it was for me to be replaced with just a click of a finger.*
>
> Fanaka, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI[20]

Informally employed workers cost less to employ per unit, receive lower wages, and limited non-wage benefits, including paid leave and social security. They can also be hired and fired in line with firm needs. These terms of employment leave workers in informal employment particularly vulnerable to exploitation and workplace violence compared to directly employed workers – especially workers hired directly by lead firms in their headquarters.

12 Scroll. Click. Suffer
    The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

Table 5: Informal hiring practices that exclude data workers from labour rights protections

INFORMAL HIRING PRACTICES		
No contracts	Task-based hiring and wages	Hiring through contractors, third parties, or manpower agencies
Short-term contracts		Hiring on a per-project basis for an hourly rate

BARRIERS TO ACCOUNTABILITY

> *I have seen others who tried to voice work-related concerns get fired. So, I have resorted to being silent.*
>
> Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI[21]

Barriers to accountability for the violence inflicted on data workers is rooted in industrial secrecy and discipline practices – including denial of freedom of association, non-disclosure agreements (NDAs), precarious work arrangements, retaliatory termination, and industry-wide blacklisting of workers who speak out.

Denial of freedom of association

> *We had workers who were threatened and others sacked for trying to unionise.*
>
> – Rahi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI[22]

Denial of freedom of association undermines the ability of workers to take collective action to intervene in structural violence and to advance, defend, and enforce all other labour rights. Data workers in Colombia, Kenya, and The Philippines told Equidem researchers that they faced retaliation for organising. In 2023, 43 workers in Kenya filed a lawsuit against Meta for unlawful redundancy. The workers said they lost their jobs with Sama, a Kenya-based firm contracted to moderate Facebook content, for organising a union. They also said they were blacklisted from applying for the same roles at another outsourcing firm, Majorel (subsequently acquired by Teleperformance), after Meta switched contractors.[23]

Non-disclosure agreements, heightened trauma, and retaliation for speaking out

NDAs, an industry norm for data labellers and content moderators, prevent workers from speaking to anyone about their working conditions. While NDAs are often justified as protecting company confidentiality, in practice, they function as a tool for silencing workers – blocking workers from reporting harm, seeking accountability, or even describing the trauma they endure on the job. This creates a dangerous dynamic: data workers experience real harms but are legally and professionally prohibited from naming or discussing them. This isolation can worsen mental health outcomes and prevent workers from accessing support or solidarity. Onyango, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, described being unable to speak about his experiences at work – even to his family: *I cannot share anything with my family. They are the ones who are the closest to me, yet I cannot share anything because I have signed an NDA. I have kept everything bottled inside of me.*[24]

Suppressing emotions related to trauma has been linked to increased stress, anxiety, depression, and even physical health issues. For instance, research indicates that emotional suppression may convey risks for earlier death. Therefore, creating environments where individuals can openly discuss their traumatic experiences is crucial for mental and physical well-being.[25] Worse, NDAs often come together with fear of retaliation – being fired, blacklisted, or sued – for speaking out. Terminating workers who raise concerns not only punishes those brave enough to speak up, but also reinforces a culture of silence, sustained by fear of retaliation. This keeps systemic exploitation hidden, allows harmful practices to continue unchecked, and erodes any path to justice or reform. In this context, NDAs don't just protect intellectual property – they protect institutionalised harm.

Ineffective grievance channels

Data workers subcontracted by BPO firms for Meta in Colombia, Ghana, and Kenya all described ineffective grievance and dispute resolution channels. Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, explained: *There are no clear dispute resolution mechanisms with management. The only available space is talking to the supervisor, which has not been effective.*[26]

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

13

LEAD FIRM FAILURE TO ESTABLISH AND IMPLEMENT MEANINGFUL CORPORATE HUMAN AND LABOUR RIGHTS COMMITMENTS

Meta and OpenAI have established formal policies outlining their commitments to human and labour rights. As laid out in Part 5 of this report on barriers to accountability, however, Equidem documented rampant violations of the Meta and OpenAI human and labour rights policies that make hollow commitments without binding channels for enforcement.

Notably, ByteDance and Remotasks lack publicly available information detailing any public commitments in these areas.

LEGAL VIOLATIONS

The working conditions faced by data labellers and content moderators – essential workers tasked with protecting a global community of social media users from harmful content – are not only abhorrent, but also illegal. The occupational safety and health harms, workplace violence, and exploitative working conditions sustained by lead firms at the helm of global content moderation and data labelling supply chains violate ILO standards, OECD Guidelines for Multinational Enterprises on Responsible Business Conduct (OECD Guidelines), the German Supply Chain Act, and the Directive (EU) 2024/1760 of the European Parliament and of the Council of 13 June 2024 on corporate sustainability due diligence (CSDDD), a multistakeholder negotiated framework for due diligence that is in the process of being codified and implemented at national levels in the EU.

The slate of violations documented by Equidem investigators is also inconsistent with the responsibilities of businesses to always respect all human rights as set out under the UNGPs, an internationally accepted set of standards endorsed by the UN Human Rights Council that extends responsibility to respect human rights beyond a company's own activities to its business relationships, such as relationships with sub-contractors. Work conditions and practices have also been challenged under the national courts in several countries, including Colombia, France, Kenya, and Spain as detailed below.

Table 6: Violations of ILO and human rights due diligence standards

	Violations of ILO standards	Violations of German Supply Chain Act	Violations of OECD Guidelines	CSDDD violations
Target pressure and industrial discipline practices to drive accelerated work				
Unreasonable targets leading to high stress and extended working hours	ILO Violence and Harassment Convention (No. 190), Article 1(a) ILO Hours of Work (Industry) Convention, 1919 (No. 1), Article 2 ILO Occupational Health and Safety Convention, 1981 (No. 155), Article 3(e), 5(a),(b) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 10(f), (g)	Section 2(5) (a),(c)	Chapter V, Section (1)(f)	Article 46, Annex, Part 1, 6
Threats of termination for failure to meet targets	ILO Violence and Harassment Convention (No. 190), Article 1(a)	Section 2(5) (a),(c)	Chapter V, Section 4(b)	Annex, Part 1, 6
Extended hours and overwork	ILO Hours of Work (Industry) Convention, 1919 (No. 1) Article 2 ILO Occupational Health and Safety Convention, 1981 (No. 155), Article 3(e), 5(b) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 3(e), 10(f), (g)	2(5)(a),(c)	Chapter V, Section (1) (f)	Annex, Part 1, 6

14   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

Forced unpaid overtime	ILO Hours of Work (Industry) Convention, 1919 (No. 1), Article 6(2) ILO Occupational Health and Safety Convention, 1981 (No. 155), Article 3(e), 5(b) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 10(f), (g)	Section 2(5) (a),(c), 2(8)	Chapter V, Section (1)(f), Section 4(b)	Annex, Part 1, 6
Team performance targets leading to workplace bullying	Section 2(5)(a),(c)	Section 2(5) (a),(c)	Chapter V, Section 4(b)	Annex, Part 1, 6
Physical Harms				
Physical impacts of exposure to violent content as a feature of work processes	ILO Occupational Health and Safety Convention, 1981 (No. 155), Article 3(e), 5(a) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 3(e), 10 (f), (g)	Section 2(5) (a),(c)	Chapter V, Section (1)(f), Section 4(b)	Annex, Part 1, 6
Extended hours and overwork	ILO Hours of Work (Industry) Convention, 1919 (No. 1), Article 2 ILO Occupational Health and Safety Convention, 1981 (No. 155), Article 3(e), 5 (b) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 3(e), 10(f), (g)	Section 2(5) (a),(c)	Chapter V, Section (1)(f), Section 4(b)	Annex, Part 1, 6
Denial of leave	ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 10(g)	Section 2(5) (c)	Chapter V, Section 4(b)	Annex, Part 1, 6
Denial of sick leave	ILO Medical Care and Sickness Benefits Convention, 1969 (No. 130) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 10 (g)	Section 2(5) (c)	Chapter V, Section 4(b)	Annex, Part 1, 6
Psychological harms				
Ongoing exposure to violent and sexually explicit content as a feature of work processes	ILO Violence and Harassment Convention (No. 190), Article 1(a) ILO Occupational Health and Safety Convention, 1981 (No. 155), Article 3(e), 5(a), (b) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 3(e), 10(f), (g)	Section 2(5) (a),(c)	Chapter V, Section (1)(f), Section 4(b), Section 59	Annex, Part 1, 6

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

15

Extended working hours conducting isolated repetitive tasks as a feature of work processes	ILO Occupational Health and Safety Convention, 1981 (No. 155), Article 3(e), 5(a),(b) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 3(e), 10(f),(g) ILO Violence and Harassment Convention (No. 190), Article 1(a)	Section 2(5) (a),(c)	Chapter V, Section (1)(f), Section 4(b), Section 59	Annex, Part 1, 6
Stress and anxiety from decision fatigue as a feature of work processes	ILO Occupational Health and Safety Convention, 1981 (No. 155), Article 3(e), 5(a),(b) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 3(e), 10(f),(g) ILO Violence and Harassment Convention (No. 190), Article 1(a)	Section 2(5) (a),(c)	Chapter V, Section (1)(f), Section 4(b), Section 59	Annex, Part 1, 6
Verbal abuse and workplace bullying	ILO Violence and Harassment Convention (No. 190) (Article 1(a))	2(5)(a)	Chapter V, Section 4(b), Section 59	Annex, Part 1, 6
Sexual harms				
Exposure to pornographic and violent sexual content resulting in sexual trauma as a feature of work processes	ILO Occupational Health and Safety Convention, 1981 (No. 155), Article 3(e), 5(a),(b) ILO Occupational Safety and Health Recommendation, 1981 (No. 164), Article 3(e), 10(f), (g) ILO Violence and Harassment Convention (No. 190), Article 1(a)	2(5)(a),(c)	Chapter V, Section (1)(f), Section 4(b), Section 59	Annex, Part 1, 6
Sexual harassment	ILO Violence and Harassment Convention (No. 190) (Article 1(a))	2(5)(a)	Chapter V, Section 59	Annex, Part 1, 6
Inadequate psychological support				
Failure to provide occupational health services, including psychological support	ILO Occupational Health Services Convention, 1985 (No. 161), Preamble, Article 5	Section 2(5) (c)	Chapter V, Section (1)(f)	Annex, Part 1, 6
Economic harms				
Unpaid overtime	ILO Hours of Work (Industry) Convention, 1919 (No. 1)(Article 6(2))	Section 2(8)	Chapter V, Section 4(b), Section 59	Annex, Part 1, 6
Task-based low wages	ILO Protection of Wages Convention, 1949 (No. 95), Article 2	Section 2(5) (c), (8)	Chapter V, Section 4(b)	Annex, Part 1, 6
Firm relocation without compensation	ILO Protection of Wages Convention, 1949 (No. 95) (Article 12) ILO Convention No. 158, Termination of Employment Convention, 1982 (Article 12)	Section 2(8)	Chapter V, Section 6	Annex, Part 1, 6

16   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

Freedom of association violations				
Retaliation for organising a union	Freedom of Association and Protection of the Rights to Organise Convention No. 87, Article 2; and 98, Article 1	Section 2(6) (a-c)	Chapter V, Section 1(a),(b), Section 59	Annex, Part 1, 13

SELECTED LISTING OF LEGAL ACTION AGAINST LEAD FIRMS IN THE NEWS

LEGAL ACTION IN KENYA
"PTSD, depression and anxiety: why former Facebook moderators in Kenya are taking legal action"
Published: December 18, 2024
Over 100 former Facebook content moderators in Kenya have filed a lawsuit against Meta and its subcontractor, Sama, alleging severe mental health issues such as PTSD and depression due to exposure to graphic content. They claim inadequate support and poor working conditions.

Court Ruling on Jurisdiction: In September 2024, a Kenyan court ruled that Meta could be sued in Kenya over the mass dismissal of content moderators by its contractor, Sama. The moderators, who were from various African countries, alleged exploitation and inadequate mental health support.

LEGAL ACTION IN SPAIN
Spanish Court Recognises Work-Related Trauma
Published: January 19, 2024
In January 2024, a Spanish judge ruled that a content moderator's trauma was a work-related injury. The moderator, employed by a Meta subcontractor, developed psychological issues after being exposed to disturbing content. This ruling was significant as it recognised the mental health risks associated with content moderation.

LEGAL ACTION IN FRANCE
Class Action Lawsuit in France
Published: June, 2023
In June 2023, a class action lawsuit was filed against Teleperformance by its shareholders, alleging that the company's growth was partly dependent on workers engaging in "inappropriate, traumatic, abusive, and potentially criminal activities." The lawsuit claimed that Teleperformance misled investors about the working conditions of its content moderators, who were subjected to widespread occupational trauma without adequate psychological support, extensive surveillance, and aggressive union-busting tactics. This legal action was significantly informed by investigative reports detailing the experiences of content moderators in Colombia.

LEGAL ACTION IN COLOMBIA
Lawsuit in colombia
Published: November, 2022
In November 2022, Colombia's Ministry of Labor launched an investigation into Teleperformance following reports of traumatic working conditions, union-busting, and low pay. Content moderators for TikTok, employed by Teleperformance in Colombia, reported earning as little as $10 per day while being exposed to distressing content, including videos depicting murder, child abuse, and sexual violence. Attempts to unionise were allegedly met with intimidation and threats. The investigation aimed to address these allegations and ensure compliance with labour standards

OECD COMPLAINT REGARDING GLOBAL PRACTICES
Allegations of Unsafe Working Conditions
In April 2020, labour unions filed a complaint against Teleperformance, a multinational business process outsourcing company, at the French National Action Point to the OECD. The complaint detailed unsafe working conditions, excessive monitoring, and union-busting activities in ten countries, including The Philippines, Colombia, and India.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

17

RESPONSES FROM LEAD FIRMS

Equidem shared all the findings of its investigations with the four companies named in this report several weeks before its publication: Bytedance, Meta, OpenAI, and Remotasks. Despite several communications with its representatives, Bytedance did not respond to requests for comment. Equidem did receive responses from OpenAI and a representative of Remotasks which are included in full as annexures to this report. A Meta representative wrote:

> *We appreciate the opportunity to respond to the concerns you have raised in the letter ahead of your forthcoming report.*
>
> *We take the support of content reviewers seriously which is why we have clear contracts with each of the companies that help review content on Facebook and Instagram that detail our expectations in a number of areas, including counselling, training and other support.*
>
> *We require all of the companies we work with to provide 24/7 on-site support with trained practitioners, an on-call service, and access to private healthcare from the first day of employment. They are also contractually obliged to pay their employees who review content on Facebook and Instagram above the industry standard in the markets they operate.*
>
> *We respect the right of our vendor's employees to organize and do not oppose or inhibit their right to unionize.*

Meta did not respond to any of the specific cases or practices we shared with them, nor provide any information about what actions, if any, they will take to investigate them further.

In a written response to our findings, an OpenAI representative said:

> *At OpenAI, we are committed to creating a safe and positive working environment for everyone involved in building our technology, including those employed by our suppliers.*
>
> *As such, I wanted to share our supplier code of conduct, which sets out clear expectations for the ethical and fair treatment of all workers, whether directly employed or part of an outsourced team. This code covers critical areas like workplace safety, fair compensation, non-discrimination, and respect for workers' rights.*
>
> *Additionally, since May 2024, our major outsource suppliers have signed up to an additional worker conditions addendum, which reinforces our expectations of partners to be aligned with our high standards around worker conditions, pay and support. We are in the process of rolling this out more broadly across our supplier network.*
>
> *After our call, we carefully reviewed the limited information you were able to provide. Unfortunately, as I flagged may be the case, the details shared were not sufficient for us to fully investigate the claims. Despite this, we conducted an investigation based on the available information and following inquiries with our suppliers, we found no evidence to support the claims.*

The OpenAI representative noted the company's interest in conducting a further investigation. Equidem is in the process of assessing whether it is possible to provide more information with the consent of the individuals interviewed without risking their safety or their jobs.

A representative of Remotasks provided a statement listing the company's commitment to fair wages, labour protections, safeguarding and support to data workers which includes "a 24/7 anonymous hotline," and mechanisms for workers to lodge complaints. The representative also noted a series of "Ongoing Improvements" they claim the company has undertaken based on worker feedback. These changes include:

- *A new, more transparent and rewards-based pay structure*
- *A consolidated platform to make it easier for contributors to ask questions, find support, and communicate about their work.*
- *Improved access to tasks, creating a more transparent experience, and supporting fairer task distribution.*
- *A new Account Support function with clearer deactivation guidelines and a simplified appeals process, and increased transparency around work allocation.*
- *Updated Terms to provide a 30-day advance notice of contract changes and clarified that task participation is voluntary and does not impact contributor ratings.*
- *Updated Community Guidelines to include a zero-tolerance policy for harassment, affirm the voluntary nature of task participation, and provide more accessible pathways for account deactivation support.*

SUMMARY OF RECOMMENDATIONS

This report exposes a global digital labour system built on institutionalised harm. Far from the sleek branding of AI as clean, contactless, and futuristic, the reality is one of deeply human labour – exhausting, precarious, and often traumatising. Behind every AI product and moderated post is a data worker facing the brunt of toxic content, algorithmic pressure, and corporate opacity. Addressing this crisis requires more than technical tweaks or marginal reforms. It demands a structural shift in how digital labour is governed, who bears responsibility, and what protections are guaranteed.

We are at a critical moment. As AI becomes more central to society, so too does the labour that sustains it. The choice ahead is stark: entrench a digital economy that profits from invisible suffering or build one grounded in safety, dignity, and justice for those who power it. This report makes clear which path we must take – and what must happen next to get there.

It proposes a structured roadmap to reduce harm and reform the conditions of data workers, presented in the final section. The recommendations are categorised by urgency

- **Immediate (0–6 months)**
- **Short-to medium-term (6–18 months)**
- **Long-term structural change (18+ months)**

They are directed at six key actors:

1 LEAD FIRMS	**2** INVESTORS	**3** THE ILO
4 HEADQUARTER ECONOMIES	**5** STATES WITH DATA WORKFORCES	**6** AI RESEARCHERS

Immediate and Actionable Measures (0–6 Months)

Urgent interventions to reduce ongoing harm and protect workers:

Lead firms should uphold freedom of association and collective bargaining rights, prohibit team-based incentive and punishment systems, ensure access to external trauma-informed mental health care, suspend the use of gag clauses in employment contracts, and mandate written contracts and paid training for all workers.

Investors should require human rights due diligence (HRDD) before investing in AI or platform companies.

Headquarter economies must mandate transparency in global outsourcing chains, including BPO disclosures and working conditions.

AI researchers should establish ethical review protocols for labour conditions, co-develop solutions with affected workers, and reject datasets built on exploitative labour.

Short to Medium-Term Reforms (6–18 Months)

Systemic policy and regulatory reforms to fix structural gaps:

Lead firms should adopt long-term, stable contracting relationships to ensure decent wages, safe working conditions, and continuity of worker protections across the supply chain.

Investors must demand compliance with labour laws and ILO standards across digital supply chains.

The ILO should formally recognise content moderation and data annotation as high-risk essential labour and set binding international standards for data work.

Headquarter economies must bar public procurement from firms that rely on exploitative outsourcing and introduce joint employer liability laws at national and regional levels.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

19
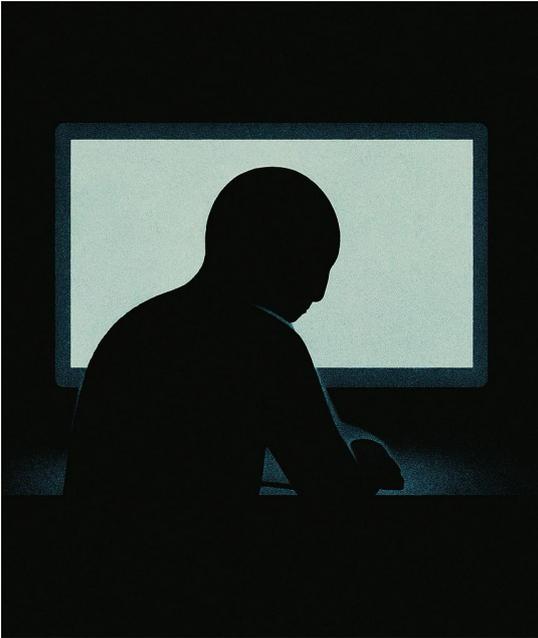
Long-Term Structural Changes (18+ Months)

Transformative shifts to rebuild data work around equity and care:

Lead firms must redesign platforms to reduce human exposure and create a global recovery and reparations fund for workers suffering long-term harm.

Investors should divest from abusive companies and adopt "wellbeing-first" funding models that centre around worker health.

AI researchers and institutions must embed labour rights and ethics into computer science and AI curricula and build public, non-extractive data labelling platforms.

This roadmap places the rights and health of data workers at the centre of AI's global infrastructure and calls for shared accountability across the private sector, states, and civil society. The recommendations presented above are detailed in the concluding chapter of this report.

20   Scroll. Click. Suffer
      The Hidden Human Cost of Content Moderation & Data Labelling

      equidem.org

METHODOLOGY

This report is based on an investigation into the experiences of workers on digital technology supply chains conducted by Equidem researchers from August to October 2024 in Colombia, Ghana, Kenya, and The Philippines. Equidem researchers engaged with 113 workers through in-person and on-line interviews, including 42 working for BPO firms subcontracted by Meta, 17 working for BPO firms subcontracted by OpenAI, and 24 for BPO firms subcontracted by Bytedance. The remaining workers were employed by BPO firms for other digital technology companies. To protect workers from retaliation, all names in this report are pseudonyms.

Table 7: Distribution of workers engaged by nationality and role

	Colombia	Ghana	Kenya	The Philippines	Total
AI Worker	3	-	5	6	14
Content Moderator	22	4	11	25	62
Content Analyst	-	1	-	-	1
Data Labeller	-	-	20	5	25
Data Annotator	-	-	2	-	2
Data Labeller Trainer	-	-	1	-	1
Quality Analyst	1	-	2	-	3
Trasncriptionist	-	-	1	-	1
Key Informant	1	2	-	1	4
Total	27	7	42	37	113

Data workers operate under intense surveillance and strict contractual obligations, particularly through the widespread use of Non-Disclosure Agreements (NDAs). During our research, many workers expressed fear of retaliation, including dismissal or legal action, which directly constrained both the depth and breadth of testimonies we were able to collect. These structural barriers are an inherent feature of the industry and not incidental, shaping the scope of this study.

- In **Colombia**, investigators reached out to 105 data workers. 75 Data workers refused to engage entirely. 30 workers agreed to engage in interviews.
- In **Kenya**, investigators reached out to 110 data workers. 68 workers refused to engage entirely. 42 workers agreed to engage in interviews.
- In The **Philippines**, researchers reached out to 54 data workers. 18 data workers refused to engage entirely, citing NDAs. 36 workers agreed to engage in interviews, and all of them mentioned signing NDAs as grounds for refusing to answer certain lines of questioning.

While our sample is not statistically representative of the global data enrichment workforce, it reflects a depth of insight into otherwise concealed working conditions. By triangulating multiple testimonies from different roles, platforms, and geographies, this report presents a qualitative snapshot of systemic harms. The patterns observed are indicative of entrenched risks faced by workers, even if they cannot be exhaustively quantified.

Worker interviews were carried out on a one-on-one basis in person and remotely through phone calls and chat messaging. Our interviews included extended structured interviews and shorter unstructured interviews. This approach allowed us to engage with the maximum number of data workers, using a detailed questionnaire where possible but also incorporating unstructured testimony when time and access constraints prohibited extended engagement.

All interviews were conducted with the informed consent of the participants, with respect for confidentiality and data security in line with Equidem's duty of care policy and procedures. Almost all the workers we interviewed requested that their identity not be revealed. Accordingly, we have used pseudonyms to shield workers from the risk of retaliation from employers. Notably, NDAs are an industry norm, making digital technology supply chain workers more hesitant to speak out about working conditions or to organise.

We contextualised these interviews through extensive desk research, including independent reporting, nascent academic literatures, and open-source company investigations through company websites and industry analysis portals.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

21

PART I
DATA WORKERS IN
THE DIGITAL PLATFORM ECONOMY

LEAD FIRM POWER IN THE DIGITAL PLATFORM ECONOMY

Digital technology supply chains – including the global technology companies behind social media platforms– are characterised by unequal relationships of power between lead firms, multinational digital solutions firms (MNEs) and third-party business process outsourcing (BPO) companies, labour contractors, and workers.[27] The global technology companies behind social media platforms hold disproportionate bargaining power in a buyer-driven market. Accordingly, these companies can source from a myriad suppliers who are essentially captive to one or a handful of buyers, a market structure that is known as monopsony.[28] This bargaining power allows global technology companies to set the terms of business relationships across their supply chains. Lead firm demands for content moderation, data labelling, and other AI work at a certain price and pace drives accelerated work for extended hours – conditions that are incompatible with decent work and human rights due diligence standards.

Figure 1: Overview of business and employment relationships between social media platforms and their supply chains

22   **Scroll. Click. Suffer**
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org
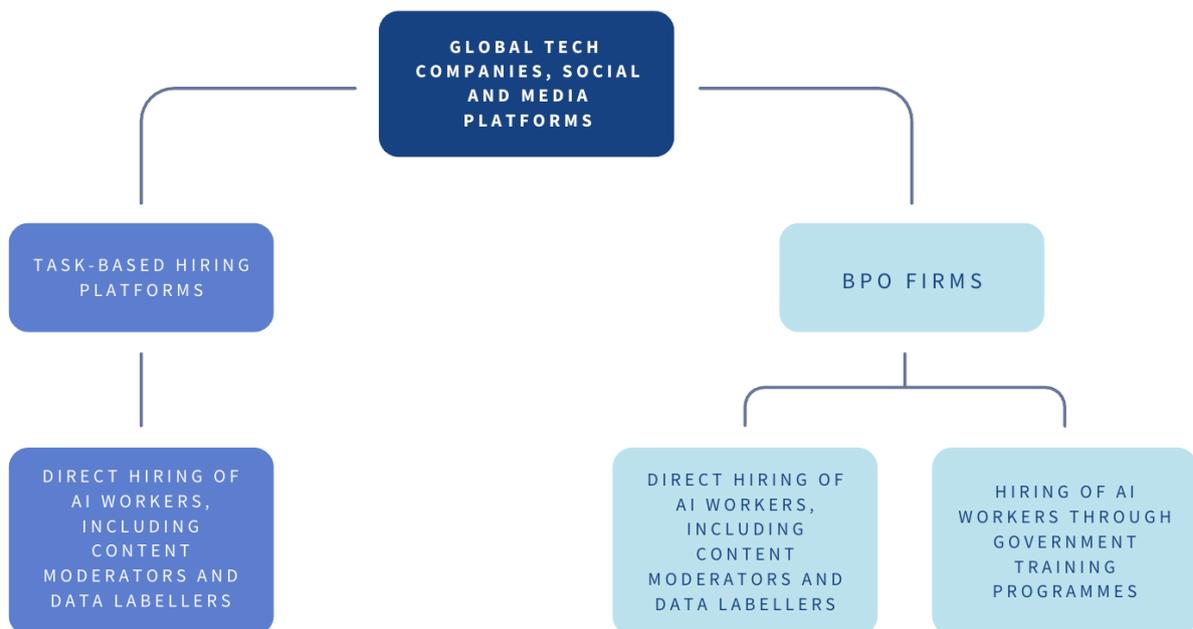
DATA LABELLING AND CONTENT MODERATION SUPPLY CHAINS IN COLOMBIA, KENYA, AND THE PHILIPPINES

Colombia, Kenya, and The Philippines have emerged as BPO hubs, leveraging their large English-speaking populations, relatively low labour costs, and favourable business environments. These countries host numerous content moderation MNEs and BPOs that employ data workers to review and manage online content for global tech giants. The Philippines is particularly prominent, with its capital, Manila, hosting many major content moderation firms, while Colombia's Bogotá and Kenya's Nairobi are also growing destinations.

COLOMBIA

Colombia has emerged as a powerhouse in business process outsourcing (BPO), ranking among the top five most preferred destinations worldwide, as confirmed by the Offshore BPO Confidence Index 2023.[29] The country's Ministry of Commerce, Industry, and Tourism reports that BPO and knowledge process outsourcing (KPO) services contribute significantly to the national economy, accounting for 3.5% of the GDP and providing employment opportunities for over 700,000 people.[30] Colombia's capital, Bogotá, has emerged as a hub for content moderation, with major companies like Accenture and Teleperformance operating in the city.[31]

KENYA

Kenya is considered a strategic hub for the content moderation industry, attracting major technology companies and startups alike, with digital technology supply chain workers reviewing content for prominent platforms and products including TikTok, Facebook, and ChatGPT.[32] Kenya's digital technology industry is anchored in Nairobi's iHub and Kilimani's Silicon Savannah, with a concentration of platform moderation industry players in Nairobi.

To address the region's linguistic diversity, moderators are recruited from countries including Burundi, Eritrea, Eswatini, Ethiopia, Nigeria, Somalia, South Africa, Uganda, Zimbabwe, and others.[33] Their crucial tasks include identifying and removing harmful content, particularly content promoted by extremist groups from their native countries, requiring nuanced cultural and linguistic understanding.

THE PHILIPPINES

The Philippines, with Manila serving as the epicentre, is a digital technology supply chain hub, employing 3-4 million workers. It accounts for 10-15% of the global BPO market, with a strong focus on social media management and content moderation. This concentration can be attributed to the country's unique advantages, including a young, educated, and English-proficient population as well as a business-friendly environment. The Philippines government and industry leaders have implemented strategic measures to foster growth in the BPO sector with investments in infrastructure, workforce training, and business-friendly policies.[34]

While the largest BPO sites are within Metro Manila, in the last decade, many BPO sites have also opened in provinces or cities from north Luzon, to Visayas and Mindanao, engaging previously unengaged semi-skilled workers in data enrichment work. Further, some BPO companies retained the work-from-home setup after the pandemic as this meant lower overheads for these companies.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

23

PART II
INFORMALITY AND EXPLOITATION OF DATA WORKERS

> *My friend was punished because she was out of her working station. She was in such a bad mental state that she did not fight back. She had a mental breakdown and left crying. She did not take a break or anything – she just had to leave, and she did. She took two hours crying with the psychologist, and when she returned, she had an action form waiting for her. After that, she just quit.*

Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance[35]

As explained by Alarico, working with no job security or employer commitment to worker well-being heightens the stress associated with his job. In data enrichment work, to meet targets at the lowest possible cost, content moderators and data labellers are hired informally. This means that they are often hired without contracts, on short-term contracts, through contractors, as freelance workers, or on a per-task and per-project basis. These terms of employment leave workers particularly vulnerable to exploitation when compared to directly employed workers.

INFORMALITY AND EXPLOITATION IN PRACTICE
REMOTASKS, KENYA

Remotasks is a global hiring platform that facilitates data annotation by workers who use the platform to earn money on a task-by-task basis. Remotasks supports the data annotation function of Scale AI, an AI company headquartered in San Francisco, California, that provides labelled data used to train AI applications. According to Scale AI, their clients include companies like Microsoft and Meta; enterprises like Fox and Accenture; generative AI companies like OpenAI and Cohere; and US government agencies like the US Army and the US Airforce.[36]

Equidem's investigation of Remotasks operations in Kenya illustrates how downstream data work suppliers implement informal employment practices that expose workers to precarity, excessive demands, and harm. Hired without contracts and paid on a per-task basis, workers lack job security, social protection, and access to effective grievance mechanisms.

Remotasks workers reported abrupt dismissal without notice, reflecting the extreme vulnerability of informally employed labour:

> *We had a crazy target to meet. If you don't meet the target, they can send you an email and sack you without any notice.*

Elinah, data labelling trainer, Remotasks, Kenya[37]

Opaque and shifting performance demands were imposed without consultation or prior disclosure, leaving workers with no choice but to accept unpaid overtime:

> *The company said that the client's targets were not met and forced us to work overtime to meet them. They did not disclose what the target was before. They just said we had to work overtime, and we had to do so.*

Dhania, data labeller, Remotasks, Kenya[38]

24   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

Exposure to violent and sexually explicit content was routine, with no access to psychological support. This resulted in serious impacts on workers' mental health and personal relationships:

> *Continuous exposure to sexual violence changed my behaviour. I feel like a different person now. It affected my relationship and ultimately led to separation.*

Linet, data labeller, Remotasks, Kenya[39]

Attempts to speak out about working conditions were met with swift retaliation:

> *I was voicing complaints on behalf of the taskers, which my team lead did not appreciate. She collaborated with the country manager, and I was fired.*

Elinah, former data labelling trainer, Remotasks, Kenya[40]

These testimonies reveal how informality is not incidental but foundational to the business model. In meeting the cost and speed demands of lead firms, Remotasks and similar suppliers displace risk and instability onto workers through practices that facilitate exploitation and shield employers from accountability.

LEAD FIRM POWER DRIVES INFORMALITY

As explained in Part 1, lead firms in the digital platform economy set the terms of business relationships across their supply chains. They can demand content moderation, data labelling, and other AI work for low wages at a high pace – conditions that are incompatible with decent work standards. The demand from lead firms for accelerated work at pay below living wages drives informalisation by incentivising downstream labour suppliers to hire workers in employment relationships that are not subject to labour and social protection standards.

Further, lead firm purchasing practices among social media MNEs are characterised by unstable relationships with downstream companies. Labour supply chains are not vertically integrated, and this means that BPO firms must bid for contracts and their renewal. In short, a BPO company holding a contract with a social media MNE can be replaced at any time.

These purchasing practices create real pressure on BPO companies lower down the chain, requiring them to quickly turn around assignments without being assured of subsequent assignments. Unstable relationships between the lead firms and downstream labour suppliers translate into unstable relationships between labour suppliers and workers. The downstream companies hire workers in informal employment to meet demand rather than as permanent workers. In short, lead firms displace market instability onto data labellers and content moderators.

Understanding informality practices as a related "spectrum of exclusions"[41] and exposure to rights violations reveals how lead firms drive informal employment, how downstream suppliers operationalise informality, and how legal exclusion creates categories of unprotected work. The analysis that follows aims to contribute to a better understanding of the dynamics of informalisation impacting content moderators and data labellers. It also seeks to advance action to address informality by providing a framework that can be used by workers, trade unions, and labour rights activists to identify informalisation practices within their workplaces and communities that expose workers to exploitation and violence.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

25

INFORMAL HIRING BY DOWNSTREAM COMPANIES

To manage the downward pressure that they face from lead firms to meet high-speed turnaround times at reduced costs, downstream companies on global data enrichment supply chains displace this pressure onto their workforce. Informally employed workers cost less to employ per unit, receive lower wages, and limited non-wage benefits, including paid leave and social security. They can also be hired and fired in line with firm needs. These terms of employment leave workers in informal employment particularly vulnerable to exploitation compared to directly employed workers.

Downstream companies on data work supply chains use a range of workforce informalisation practices (Table 8). Informal hiring practices have significant economic consequences for digital technology supply chain workers who experience unpredictable wages and work under the constant threat of termination.

Table 8: Informal hiring practices documented by Equidem on digital technology supply chains in Colombia, Kenya, and The Philippines

INFORMAL HIRING PRACTICES		
Employing Workers Without Contracts	Employing Workers On Short-Term Contracts	Hiring Through Contractors, Third Parties, Or Manpower Agencies
Task-based hiring and wages	Hiring on a per-project basis	Freelance work

The time for short-term contracts, moreover, is deliberately set to exclude workers from protections. For instance, under Kenya's Employment Act, 2007, after six months, workers are eligible for regular employment and associated benefits. The practice of hiring data workers on fixed-duration contracts for less than six months systematically excludes them from labour rights and social protection.

26   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

PART III
TARGET PRESSURE AND INDUSTRIAL DISCIPLINE PRACTICES

TARGET PRESSURE

Content moderators and data labellers respond to flagged content with decisions on whether to delete the content or on how to label the content, respectively. They work through a queue of images throughout the day. Workers are evaluated based on productivity (the number of cases they get through) and accuracy (whether a supervisor reviewing cases would agree with their decisions).[42] **Twenty-six workers from Colombia, Ghana, Kenya, and The Philippines described being driven to reach unmanageable targets.**

Bitok, working as a content moderator for a BPO firm in Kenya subcontracted by ByteDance, linked target pressure to project deadlines set by lead firms: ***I feel pressure on certain projects because some have strict deadlines that have to be met within a specific time frame.[43]***

Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, described intense pressure and unrealistic expectations at work: ***It's a heavy job by itself. It doesn't need the pressure of a call centre to make it worse. There are also other things like AHT (average handle time), where you have to do one job every 90 seconds. It's insane to ask this from the worker in this line of work. They expected more quantity than quality. You can't have both every time.[44]***
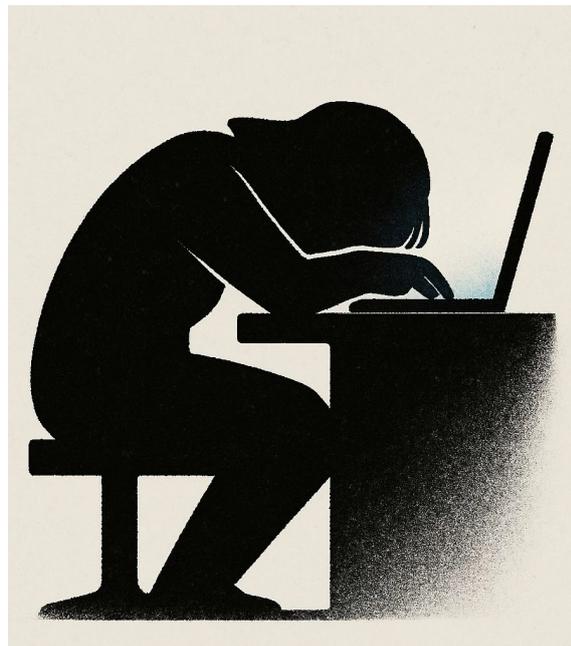
Mile, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, reported having less than 10 seconds per image in a queue that could run up to 700-1,000 cases: ***We have to moderate a lot of cases. In a normal shift, we had to go through 700 to 1,000 cases, which means we have to accomplish an average time of 7-12 seconds to moderate each case.[45]***

Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, described hourly targets: ***The videos are very disturbing in nature, and one has to go through 300 videos per hour.[46]***

Digital technology supply chain workers in The Philippines who work on moderating and labelling videos described having to clear over 1,000 videos in a queue. As one Filipina worker explained: ***Imagine, they expect 1,000 [videos] a day to be finished in your queue, so really you only have 10 seconds, and if you are caught taking more time more than five times in a day, or if you finish less than 5,000 per week, then you are in trouble. The standard is so high – and it's all just based on those scorecards.[47]***

Edna, working as a content moderator for a BPO sub-contracted by Meta and ByteDance, described escalation of targets over time: ***At first it was manageable, but as time went by, they added more tasks to the queues. It became unbearable.[48]***

According to Professor Sarah Roberts at UCLA, workers may not even be provided with the interfaces required to move at the speeds their employer demands. According to Roberts, moderators often use interfaces that have not been recently updated and may therefore respond slowly to a moderator's commands, both creating inefficiency and prolonging the exposure to harmful content.[49] Consistent with this account, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, described computer malfunction as routine: ***My Colombian colleagues used to reach out to me and ask me to help them when they were struggling with their computers because I was known for speaking out. 'Can you ask our supervisor to fix this computer? It's really bad and I'm struggling to work,' they would say.[50]***

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

27

Evans, a migrant worker in Ghana employed as a content moderator by a BPO firm subcontracted by Meta and ByteDance, described heightened stress associated with platform shutdowns: ***The worst thing about this job is the platform we are working on. I think it is affecting my mental health. The system automatically shuts your screen and puts your status to unavailable if you don't touch the mouse or keyboard, even if you don't have a ticket to work on. It's a bad experience. It is so stressful.[51]***

INDUSTRIAL DISCIPLINE PRACTICES USED TO DRIVE ACCELERATED WORK

The digital technology supply chain workers interviewed by Equidem researchers described being pushed to meet unreasonable targets, driven by fear of severe penalties for falling short. The penalties for failing to meet production targets include forfeiture of target-based bonuses, forced unpaid overtime, and pay cuts. Some workers also described team-based targets, which drove workplace bullying.

PAY CUTS

Nine data workers, including workers from Colombia and Kenya, told Equidem researchers that they were threatened with pay cuts if they did not meet targets, coercing them into working unpaid overtime hours.

Aziz, working as a quality analyst for a BPO firm in Kenya subcontracted by Meta and OpenAI, explained: ***The company pressurises us to meet the weekly target. We must work overtime, which is unpaid, to meet these targets. If we do not meet the target, the company cuts our pay.[52]***

Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, told Equidem that salaries were contingent upon achieving fast-paced accuracy: ***There were targets that one had to achieve to receive the entire salary-average handling time and accuracy scores (the number of times you moderate a video correctly). If both were good, then you would be on the safe side and receive your salary.[53]***

WITHHOLDING BONUS PAYMENTS

Workers also described having their bonus payments made contingent on meeting targets. As explained by the workers we spoke to, the bonus offered by companies made a significant portion of their income. In some cases, the bonuses accounted for more than 70% of workers' total remuneration. **9 workers, including workers from Colombia, Ghana, and Kenya, described manipulation of conditional bonus payments to drive production targets.**

Dumba, a migrant worker in Ghana working as a social media content moderator for a BPO firm subcontracted by Meta and ByteDance, told Equidem researchers that his bonus – which accounts for 72% of his total income – is dependent on meeting targets: ***My basic salary is 1,400 cedis (90 USD), and with the bonus, I can earn up to 5,000 cedis. However, to qualify for the bonus, I must achieve at least 85% of the set target. Unfortunately, the base salary alone is too low to meet my financial needs. Without the bonus, it's difficult to make ends meet.[54]***

Kojo, working as a content analyst for a BPO firm subcontracted by Meta and ByteDance in Ghana, told Equidem researchers that his bonus payments account for 50% of his wages, and that if

he does not meet his targets, his bonus is cut: ***My basic salary is 1,100 cedis (70 USD). I get 1,130 cedis extra in bonus. If I do not meet the target, my bonus gets slashed – that's almost equal to my basic wage.*** [55]

Evans, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, shared: ***Only 20% of the total salary is a basic wage; the rest is tied to bonuses, which are tied to targets and performance.*** [56]

Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, shared that the company penalised workers for not hitting the set target: ***If we did not meet the target, we got coaching sessions and then they deducted your bonus. If repeated, you could get an Action Form, which is basically an administrative measure saying that you did something wrong. We had 119 USD as a bonus, but it was always so difficult to get. I never really counted on them.*** [57]

According to Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, the BPO firm constantly shifted the conditions for earning performance-contingent bonuses: ***There was a bonus, but I don't think I got it more than once or twice. They constantly changed the rules for getting it. They kept changing what affected the bonus, and the weight of each thing. It was attendance, number of jobs done, number of jobs correct, and a weekly test. The target of each thing and how much they influenced your bonus changed almost monthly.*** [58]

Team performance targets and workplace bullying

In The Philippines, data labellers, content moderators, and other data enrichment workers are evaluated based upon team targets. The collective ability to meet targets determines the tenure and promotion of the team leader, and the team schedule, including holidays and weekends off. Under this system, the entire team is penalised for low performance by any team member.

The workers interviewed by Equidem researchers described significant challenges meeting team targets during their initial months on the job, both in reaching the targets on their "score cards" and in coping with the constant exposure to harmful content. The combined pressure to meet performance metrics and the emotional toll of the work creates a toxic work environment, particularly for new members to the team. Workers reported bullying from peers and team leaders as a normalised part of the work culture. As explained in the following chapter, target pressures and associated industrial discipline practices fuel workplace violence across data enrichment supply chains.

Threats of termination

Data workers also described the very real threat of termination if they did not meet targets. Fanaka, working as a content moderator for a BPO firm subcontracted by Meta and OpenAI, explained to Equidem investigators that failure to meet targets could result in probation and termination: ***The targets were crazy keeping in mind how heavy the content is. If you could not reach the target, you would be put on a performance improvement programme (PIP), which later may lead to your dismissal.*** [59]

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

29

PART IV
HEALTH AND SAFETY RISKS
AND WORKPLACE VIOLENCE

> *I left the job because I could not bear it anymore. In the past months, it got worse. I could not sleep properly and had nightmares about the job. It also produced anxiety, and I lost weight because of it. When I talked to my colleagues, they had symptoms similar to mine. This job... it messes with you.*
>
> Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance [60]

Workers on digital technology supply chains in Colombia, Kenya, and The Philippines are subjected to a slate of occupational health and safety risks, amounting to systemic workplace violence under the ILO Violence and Harassment Convention, 2019 (No. 190), which defines violence and harassment as "unacceptable behaviours and practices" that "aim at, result in, or are likely to result in physical, psychological, sexual, or economic harm."

ACTS THAT INFLICT PHYSICAL HARM

Workers on digital technology supply chains described the physical harms they faced due to sustained exposure to explicit content, overwork, and financial penalties for taking sick leave.

Table 9: Physical harms reported by data workers

PHYSICAL HARMS		
EXHAUSTION	INSOMNIA	PSORIASIS
EYE STRAIN	NIGHTMARES	SUBSTANCE DEPENDENCE
HAIR LOSS	PAIN IN THE FEET, BACK AND WRISTS	WEIGHT LOSS

These harms were widespread and widely reported: Equidem researchers recorded 31 reports of physical harms from data workers across Colombia, Ghana, and Kenya.

Table 10: Incidents of physical harms reported by data workers

	Colombia	Ghana	Kenya	Total
Physical harms	13	2	16	31
Number of Workers	27	7	42	76

Source of information: This table is based on a qualitative review of 76 worker testimonies from Colombia, Ghana, and Kenya. Each testimony was analysed to identify references to physical harm. Multiple harms described within a single testimony were individually counted. Where workers mentioned several impacts under the same category (e.g., both weight loss and psoriasis), these were recorded as separate instances. Accordingly, this data represents the total number of harms reported. This approach aims to highlight not just the presence of harm, but the intensity and prevalence of harms reported. Data from The Philippines was excluded from this table due to limited individual testimonies.

30

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

Physical impacts of exposure to violent content

> *I had to read messages between people from many parts of the world. These messages were flagged by the algorithm and people. I had to see child pornography, death threats, torture – but the majority of the cases were child pornography.*
>
> Antonio, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance [61]

Workers described the impact of sustained exposure to explicit violent and sexual content. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, highlighted that his job included moderating messenger chats on Instagram and Facebook and explained the physical impacts it had, including weight loss, psoriasis, and insomnia: *I specifically used to moderate chats for messenger on Instagram and Facebook. Content was hard, mostly child sexual exploitation. Later I was moved to IIC, which stands for inappropriate interactions with children. When I was moved to this queue, the content was 99% child pornography and abuse material. I was dealing with inappropriate content day in and day out for 8 hours a day. I lost around 15 kilos in 6-8 months; I have psoriasis, which has gotten much worse due to stress; I have had terrible sleeping patterns while working there, especially when I was working the graveyard shift. I felt more depressed, and sometimes I would catch myself drifting off and thinking of videos I had seen at work and stuck with me.* [62]

Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, described insomnia from the psychological toll of his work: *At the end of the day, one usually gets tired and prone to psychological effects. I suffer from insomnia and PTSD from the work I did.* [63]

Edna, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, described the physical toll of anxiety and nightmares: *The constant anxiety I felt at work and nightmares impacted my mental and physical health badly.* [64]

Moiben, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and Open AI, described insomnia as her new normal: *I was unstable and became numb. That somehow became my new normal. I started losing sleep. I'm used to it now.* [65]

Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, explained: *Some jobs would trigger awful responses in my body and mind and could make me feel bad for days.* [66]

Overwork

The impacts of sustained exposure to explicit content are heightened by overwork in an industry where, as described in the previous section, workers are coerced to complete queues of up to 1,000 videos per day. Long hours sitting at a computer also has physical impacts on the body. Daina, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and Open AI, described the physical impacts of her work: *I experienced pain in my feet and swelling, back pain due to long working hours, and wrist pain due to typing for long hours. Even today I have to put my legs up due to swelling. I also now use prescription glasses.* [67]

Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, described using "well-being breaks" to use the washroom since he is not allowed to leave his seat during production hours: *They have well-being breaks which you are supposed to have weekly. Around 90 minutes, but it really depends on how nice your supervisor wants to be. You're not allowed to leave your seat during production hours. So, if you need to take a break to use the bathroom, you need to use some other time you have, like meeting time or training time, but if you don't have any, you use your well-being time.* [68]

Kijani, working as a quality analyst for a BPO firm in Kenya subcontracted by Meta and OpenAI, explained that her working hours depend upon client targets: *The management claimed that the client's targets were not met. I had to work extra time to meet the targets.* [69]

These targets, moreover, are not stable. Rather, they shift in response to global events that increase violence on the platform. Moiben, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and Open AI,

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

31

explained: ***There is pressure to work overtime when content increases on the platform – especially during elections or in times when something big happens, for example an attack somewhere, like the al-Shabaab attack on Westland Dusit.***[70]

Chacha a migrant worker in Kenya working as a content moderator for a BPO firm subcontracted by Meta and OpenAI, described being made to work extended hours to cover for other workers on his team: ***I had no personal time to myself. Sometimes we would be told to extend hours because some colleagues are not present or are sick. We would also be told to work on our off days and holidays.***[71]

Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, told Equidem that he was made to either extend working hours or work during weekends to meet targets: ***I have to use my days off, including weekends or extend my working hours to meet targets.***[72]

In the most extreme cases, news media has reported death as a consequence of overwork. As reported by AP News on March 19, 2025, content moderators in Kenya held a vigil for Ladi Anzaki Olubunmi, a TikTok moderator employed by Teleperformance Kenya and subcontracted by ByteDance, who passed away after expressing fatigue and a desire to return home. This incident has raised concerns about the working conditions and mental health support for content moderators.[73]

DENIAL OF LEAVE

> *The job offer and contract said we would have 15 days of unpaid time off and 15 of paid time off. I never once got either approved. Then, I did get one approved to attend a family event and that was revoked the day before my flight. I'm guessing they did not give us leave because the company would not be able to deliver the hours the client contracted.*
>
> Alarico, a migrant worker in Colombia employed as a content moderator by a BPO [74]

Increasing the toll of this work still further, workers described being unable to take off on weekends, or secure vacation days or even sick leave. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, described being unable to get leave for a surgery: ***For five months, my supervisor tried to block my vacation and interfered with a surgery I scheduled. I managed to take only half my vacation while the remaining time from my second year was blocked due to understaffing and other employees' vacations. I requested my vacation for August in the previous December, so it's difficult to understand why they couldn't plan for it. I was repeatedly told my request was not a priority.***[75]

Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, described being unable to get leave on Saturdays or Sunday to spend time with his family: ***It was normal not to have the days off you wanted or needed. If I wanted Saturdays and Sundays off to be with my family, I probably couldn't have it. These were competitive days to have our days off and usually, people fought for them. We had a heat map which showed how many employees you needed to have in on X-day or Y-day.***[76]

Even when workers are granted sick leave, they described facing financial penalties, including loss of wages and deductions from their bonuses. Dumba, a migrant worker in Ghana employed as a social media

32   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

content moderator by a BPO firm subcontracted by Meta and ByteDance, reported being penalised for taking sick leave: *There is no real sick leave. When you are sick even for one day, you don't get the attendance bonus. The attendance bonus is 150 cedis (10 USD). In addition, 10% of all other bonuses are deducted for each day some-one is sick.[77]*

Workplace stress leading to substance abuse

Aziz, working as a quality analyst by a BPO firm in Kenya subcontracted by Meta and OpenAI, described turning to alcohol to manage the stress of his work: *I have not been able to cope with the stress that comes with seeing all this explicit content. I feel like I am becoming more and more dependent on alcohol to suppress the stress.[78]*

Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteD-ance, reported that the company expected their employees to be available 24 hours a day 7 days a week for work. She described the impact on workers' health, including disassociation and drug use: *They take the "24-hour availability" way too seriously. It is like you sold yourself to them. They change your shifts, hours, and every-thing whenever they want to. They don't want you to react, they want you to be so disassociated that you don't stress anymore. And it is good money, so I understand why people stay. I wouldn't be surprised if anyone ever jumped from [the] stairs. We hear people doing drugs in the bathroom any time of the day. It is a horrendous place to be in. I never thought mental health could get this serious.[79]*

ACTS THAT INFLICT PSYCHOLOGICAL HARM

Data workers confided in Equidem researchers about the toll that content moderation and data labelling has taken on their mental health, lives, and families. The inherent psychological harms associated with moderating violent content, moreover, are exacerbated by constant pressure to meet targets within timelines set by lead firms – re-quiring extended working hours and even 24-hour availability. As discussed in Part 3, downstream BPO companies drive workers to perform through industrial discipline practices, including threats of wage and bonus loss, termi-nation, verbal abuse, and bullying. These abusive working conditions exacerbate the mental health impacts of repeated and ongoing exposure to violent and sexually explicit content.

Table 11: Psychological harms reported by data workers

PSYCHOLOGICAL HARMS		
AGGRESSION	IRRITABILITY	SOCIAL WITHDRAWAL AND ISOLATION
ANXIETY	NIGHTMARES	STRAIN ON FAMILY RELATIONSHIPS
DECISION FATIGUE	NIGHT TERRORS	STRESS FROM TARGETS AND PENALTIES
DEPRESSION	NUMBNESS	SUBSTANCE DEPEN-DENCE
DETACHMENT AND INDIFFERENCE	PANIC ATTACKS	SUICIDAL IDEATION
EMOTIONAL FATIGUE	POST-TRAUMATIC STRESS DISORDER	TRAUMATIC MEMORY RECALL

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

33

These harms were widespread and widely reported: Equidem researchers recorded 60 reports of psychological harms from 76 data workers across Colombia, Ghana, and Kenya.

Table 12: Incidences of psychological harms reported by data workers

	Colombia	Ghana	Kenya	Total
Psychological harms	25	4	31	60
Number of Workers	27	7	42	76

Ongoing exposure to violent and sexually explicit content

> *I watched a video of a babysitter who hit and kicked a newborn baby. I will never forget that child crying in pain. When I assist or hear news about child abuse, that video comes to my mind. I will remember that forever. The second one was an ex-husband who shot his ex-wife when she arrived at her home. The screams from the woman when she recognises the guy and then he shoots are impressed on my mind.*
>
> Samuel, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance [80]

Digital technology supply chain workers confided in Equidem researchers about the types of content that impacts their mental health. Camilia, working as a content analyst by a BPO firm in Colombia subcontracted by Meta and ByteDance, described the type of content exposure that impacts her most: *It really depends on what disturbs you. For me it is child abuse and animal abuse – we can see that – and cadavers (and their violation), and human trafficking, sexual solicitation, and prostitution. All of this can be disturbing.*[81]

Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, explained: *I was exposed to things from very explicit murders and torture to suicide, child abuse – things I honestly never imagined people did, let along post online. I saw this content at least a couple of times a day.*[82]

Social isolation leading to withdrawal from family and other relationships

Data workers described the experience of sitting alone for hours in front of their computers, parsing violent imagery as socially isolating. Withdrawal from family and loved ones as a result was commonly described.

Fanaka, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI confided to Equidem researchers about the mental impact the job had on her: *It exposed me to the things I wasn't ready to see, and I was not even told about them before starting the job. I cannot socialise with my family the way I used to before the job, which makes me a different person who just isolates herself. I am anxious and feel fear. I know my life will never be the same again.*[83]

Iman, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, explained: *The stress and emotional fatigue from handling distressing content has impacted my mood and mental well-being. It has been affecting my relationship with my family.*[84]

Daudi, working as an AI worker at a BPO firm in Kenya subcontracted by Meta and OpenAI, described night terrors and talking in his sleep – signs of PTSD and extreme stress that can occur after witnessing a traumatic event: *I keep having bad dreams. My family tells me that I speak of awful things while I am asleep. They are horrified for me, even scared of me because of the things I say.*[85]

Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, described detachment and self-isolation as a mental health outcome of his work: *I feel I am getting more and more detached from my family. After doing this work, I don't feel like talking to anyone.*[86]

34   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

Verbal abuse

Miguel, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, described being subjected to verbal abuse by his supervisors: **The bosses at the operation floor say abusive and inappropriate things to us. For the clients we work for, we are not allowed to have phones or any digital equipment with us on the operation floor. This has prevented us from recording these abuses.[87]**

Evans, a migrant worker in Ghana hired as a content moderator for a BPO firm subcontracted by Meta and ByteDance, described his working environment as stressful and toxic: **With the strict micro-management rules, this environment is becoming more stressful and toxic than the actual content I'm moderating.[88]**

Decision fatigue

Not only do long queues expose workers to a tremendous volume of harmful content, but digital technology supply chain workers are also required to make decisions about this material. According to the American Medical Association, making decisions day in and day out can be exhausting and cause people to feel overwhelmed, anxious, or stressed. According to psychiatrist Dr. Lisa MacLean: **A person with decision fatigue may feel tired, have brain fog, or experience other signs and symptoms of physical or mental fatigue. The phenomenon is cumulative so that as the person makes more decisions, they may feel worse or more drained as the day progresses... If your brain is worn down, it may cause you to become more reckless with your decision making or not think things through. It can also cause you to simply do nothing, which can cause even more problems.[89]**

Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, described feeling more irritable and aggressive as a result of his work: **I could notice myself more annoyed and less patient, and for the last few months, I noticed I was having more violent tendencies. I was a very peaceful person who had never had a fight before. I'm very afraid of fighting, but I started thinking about starting fights with strangers at the smallest annoyance or inconvenience.[90]**

Risk of retaliation from content producers

Data workers explained to Equidem that processes of escalating and removing user posts on Meta platforms can present risks for themselves and their families. Evans, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, explained: **You may already be aware of the ongoing conflict in Ethiopia's southwest and north between the government and various armed groups. The company has designated these groups and their affiliates as dangerous organisations and individuals. As a result, I am responsible for removing much of their content from the platform. I'm particularly concerned because one of these armed groups operates near my father's birthplace – and they are already aware that I work for this company and that I remove their content. This means my role puts both my life and the safety of my family at serious risk.[91]**

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

35

ACTS THAT INFLICT SEXUAL HARM

Table 13: Sexual harms reported by data workers

SEXUAL HARMS		
EMPLOYER REFUSAL TO ADDRESS WORKPLACE SEXUAL HARASSMENT	SEXUAL HARASSMENT AND VIOLENCE IN THE WORKPLACE	TRAUMA FROM LABELLING AND MODERATING SEXUALLY EXPLICIT CONTENT
IMPACT ON INTIMATE RELATIONSHIPS		SEXUAL WITHDRAWAL

Table 14: Incidences of sexual harms reported by data workers

	Colombia	Ghana	Kenya	Total
Sexual harms	3	-	7	10
Number of Workers	27	7	42	76

EXPOSURE TO PORNOGRAPHIC AND VIOLENT SEXUAL CONTENT RESULTING IN SEXUAL TRAUMA

Onyango, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, described the impact of prolonged regular exposure to graphic sexual content on his ability to engage in healthy sexual relationships: *I am constantly exposed to sexually violent content. This has affected my sex life. I feel detached and no longer have the desire to have sex.*[92]

According to Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, pornography and content related to adult and child exploitation usually depicts women and children: *I often came across explicit pornography and content related to adult and child exploitation. While animal abuse was less common because the system quickly deleted those graphic images, there was a lot of misogynistic material. I saw many comments and groups targeting women and children. As a young woman, it was really upsetting to witness this.*[93]

Consistent with Sofia's account, research establishes that the vast majority of violent sexually explicit content depicts violence against women.[94] Clinical studies on the impact of violent sexual content on women found that women exposed to this content experienced increased anxiety, depression, and PTSD[95] as well as decreased self-esteem and body satisfaction.[96]

Equidem researchers also documented the failure to protect workers from sexual harassment and violence in the workplace. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, described facing sexual assault on the production floor and being unable to get relief:

> *I told them a colleague has abused me, touching me in between my legs on the production floor. I made reports with my supervisor and HR [human resources]. I sent emails and followed up.*
>
> *One year later they called many women to watch a speech about abuse in the workplace and they were telling us that "not everything is abuse and you have to know how to set a difference."*

36   Scroll. Click. Suffer
      The Hidden Human Cost of Content Moderation & Data Labelling

      equidem.org

> *I just went insane, I had lost 13-15kgs, working overnight, my hair was falling and I just lost it, I told them I couldn't believe they were teaching us what was abuse and not actual abusers what not to do. So, I mentioned my one-year-old case that they never updated me on, and the person was still working with me.*
>
> *So these women from the psychology team started talking to me again, collected all the same information that they had, and told me they were escalating it. Later, this was what the psychology head manager sent me: "you have to have a better understanding of cultural differences and be aware that someone touching your back was not sexual."*
>
> *I had a few sick days after that because I never felt that disrespected in my life. This woman changed the word I used, the only word I used, I am 100% sure I did not ever say or write "back." I know where he touched me and how I felt about that. I just couldn't believe how low she was.*
>
> *After that the same guy made more victims and they just had to fire him under "overstaffing" explanation because the girls were making a formal complaint in a group and that would affect them. But if anyone asks, it was overstaffing.[97]*

Edna, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, also described sexual harassment at work impacting numerous women, albeit with fewer details.[98]

Notably, documenting workplace sexual violence and harassment across industrial contexts is particularly challenging since workers who experience sexual harassment and violence face heightened barriers to reporting due to stigma and fear of retaliation from perpetrators. Accordingly, these reports of failure to protect workers from sexual harassment and violence in the workplace suggests the need for further investigation into this area.

ACTS THAT INFLICT ECONOMIC HARM

Workers on digital technology supply chains are exposed to a slate of practices that inflict economic harm (Table 15). These acts, in conjunction with industrial discipline practices that inflict physical, psychological, and sexual harm, heighten experiences of violence for data workers.

Table 15: Economic harms reported by data workers

ECONOMIC HARMS		
UNSTABLE EMPLOYMENT	LOW WAGES	UNFAIR WAGE DEDUCTIONS
NO FIXED SALARY	LOW TASK-BASED PAYMENTS	FINANCIAL PENALTIES FOR TAKING SICK LEAVE
FORCED UNPAID OVERTIME	FINANCIAL PENALTIES FOR NOT MEETING TARGETS	FORCED PERIODS OF UNEMPLOYMENT WITHOUT PAY
MANDATORY UNPAID TRAINING	UNFAIR DENIAL OF BONUS PAYMENTS	

Equidem researchers recorded 52 incidences of economic harms from data workers across Colombia, Ghana, and Kenya.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

37

Table 16: Incidences of economic harms reported by data workers

	Colombia	Ghana	Kenya	Total
Economic harms	24	6	22	52
Number of Workers	27	7	42	76

UNPAID OVERTIME

Iman, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, reported working extended overtime without compensation: ***There was some pressure to work overtime or come in on days off, especially during peak times or holidays. This was often due to increased workload or to ensure that all content moderation tasks were completed on time, but I did not earn overtime pay for working longer hours.[99]***

Aziz, working as a quality analyst at a BPO firm in Kenya subcontracted by Meta and OpenAI, also described unpaid overtime to meet targets: ***The company pressures us to meet the weekly target. We must work overtime which is unpaid to meet these targets. If we do not meet the target, the company cuts our pay.[100]***

TASK-BASED WAGES

Kijani, working as a quality analyst for a BPO firm in Kenya subcontracted by Meta and OpenAI, described being paid by task at a very low rate, leading to an unpredictable salary: ***Some tasks were based on the number of tasks done and the accuracy of the tasks, and the wages were very low. Our salary is calculated based on quality and utilisation of time. There is no specific salary amount.[101]***

As explained by Gladys, data labellers paid by task are put in an impossible bind – they must both speed up and ensure uncompromised accuracy to receive already low task-based wages without deductions.

FIRM RELOCATION WITHOUT COMPENSATION

Data workers told Equidem that it was common practice for global BPO firms to relocate operations when faced with legal challenges from workers. Hamidi, a Kenyan national working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, explained: ***The company I worked for closed shop in Kenya and relocated their moderation to Nigeria. There is a tendency to close shop and move to another country when they are threatened with legal battles.[102]***

Iman, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, described abruptly losing her job without compensation when Meta decided to relocate: ***There are issues of company relocation. Meta decided to move their operations from Kenya to Ghana, which led to [the BPO firm] dismissing us. We were declared redundant and lost our jobs without proper compensation.[103]***

Evans, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, explained: ***Then, on November 8, 2023, after eight months with the company, our contracts were terminated due to redundancy. They did not compensate us for the remainder of the contract and merely said the project might continue in a different country, promising to contact us if needed. So, we all returned to Addis Ababa.[104]***

38   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

FAILURE TO PROVIDE ADEQUATE PSYCHOLOGICAL SUPPORT

> *They do not care about the welfare of their employees. No psychological checks, no well-being checks, and low salaries – this does not measure up to the kind of work of content moderators have to face.*
>
> Union organiser in The Philippines, describing working conditions in a BPO firm in The Philippines subcontracted by Meta and ByteDance [105]

The mental health toll on data workers is systemic and severe. While some companies presented wellness programs on paper, in practice these were poorly communicated to workers, ineffective, and inconsistently accessible. In some cases, workers told Equidem they were punished for taking the time to use even the limited mental health resources available. While supportive supervisors can make all the difference to worker wellbeing, supervisors face retaliation for advocating on behalf of workers at the expense of lead firm production targets. These failings are driven by the extreme pressure to meet performance targets at the expense of worker wellbeing. Inability to access meaningful mental health support at work is exacerbated by private health insurance company exclusions from mental health support. For instance, while data workers in The Philippines overwhelmingly reported that health insurance is among the benefits provided for BPO workers, including content moderators and data labellers, most private health insurance companies do not cover mental health treatment.

FAILURE TO INFORM WORKERS ABOUT MENTAL HEALTH SUPPORT

Workers told Equidem that they were unaware of mental health support available in the workplace. Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, shared: ***We had a team dedicated to mental health and you could take the time you need for it at the workplace. The thing is most people didn't know that. Our supervisors never said anything about the wellness teams and how to reach out to them. My trainer told me about OHS policies but it was really glossed over so that I would not have any deep knowledge in this topic. I only know what I know because I looked it up on my own. So only a few people knew about the rules and that you would not be punished. So they would stress out at their workstations, or worse, have mental breakdowns.*** [106]

INEFFECTIVE MENTAL HEALTH SERVICES

Even when workers accessed available services, they were sometimes counterproductive. For instance, workers told Equidem that group sessions intended to help workers process trauma instead triggered anxiety and further distress. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, explained: ***We would sit in a circle and talk about traumas. That was awful. People had anxiety attacks, not because we were forced to share, but because there was always someone sharing either about their personal life or especially about content that they saw that wasn't nice, and content about the jobs we saw would trigger other people. They didn't know how to deal with crises at all.*** [107]

PENALTIES FOR USING WELLNESS BREAKS

Workers told Equidem that wellness breaks were essential to their ability to spend extended hours moderating violent and explicit content – but that they faced penalties for taking breaks in the face of unreasonable production targets.

Data workers who reported to BPO firms for work even described wellness facilities to be used during these breaks. Philip, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, explained: ***They had psychologists, game rooms, they had some hours per week for coaching or if you needed that time just for relaxing, you could go ahead.*** [108]

But workers told Equidem they were not just unaware of wellness breaks – they feared punishment for using them since BPO company policies cap allowable break times. This disconnect forces workers to choose between protecting their mental health and safeguarding their jobs. Miguel, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, explained: ***[The BPO firm] told us that using more than 1.5 hours of Wellbeing Pauses per week could result in action forms.*** [109]

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

39

Data workers also told Equidem investigators that those who took wellness breaks faced disciplinary measures. Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, shared: *My friend was punished because she was out of her working station. She was in such a bad mental state that she did not fight back. She had a mental breakdown and left crying. She did not take a break or anything – she just had to leave, and she did. She took two hours crying with the psychologist, and when she returned, she had an action form waiting for her. After that, she just quit.[109]*

Due to pressure to meet production targets, data workers also described using wellness breaks to use the bathroom since they were otherwise not allowed to leave their seats during production hours. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, explained: *They have wellbeing breaks which you are supposed to have weekly. Around 90 minutes, but it really depends on how nice your supervisor wants to be. You're not allowed to leave your seat during production hours. So if you need to take a break to use the bathroom you need to use some other 'time' you have. Meeting time or training time, but if you don't have any, you use your wellbeing time.[110]*

In these instances, the drive for speed and accuracy, combined with fear of punishment, forces workers to suppress their psychological needs in favour of unrealistic quotas. In this context, mental health supports are treated as a liability to production, rather than an essential right.

Retaliation Against Supportive Supervisors

Workers told Equidem that supportive supervisors could make significant differences in their wellbeing. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, told Equidem that supervisors determined whether workers could take wellbeing breaks: *They have wellbeing breaks which you are supposed to have weekly – around 90 minutes – but it really depends on how nice your supervisor wants to be.[111]*

Samuel, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, described receiving meaningful support from his supervisor: *I spoke with my supervisor. He enabled the protocol to redirect disturbing videos. Also, in our one-to-one sessions, they would ask questions like, 'how do you feel? Do you need psychological assistance?' He tried to keep the safest work environment possible. Of course, my supervisor was very approachable and open to giving us secured and private spaces to talk about our mental health.[112]*

Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, explained that having had both bad and good supervisors, good supervisors could make all the difference to worker wellbeing: *I had bad and good managers. The good ones do wonders for your mental health in this area; it's really important.[113]*

Supervisors, however, told Equidem that they faced retaliation for advocating for workers' mental health needs. In some instances, supportive managers were penalised, or even terminated, for prioritising the well-being of their teams over relentless productivity targets. Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, told Equidem that as a manager he tried to advocate for his colleagues but was fired and blacklisted for speaking out: *I used my manager's*

40   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org
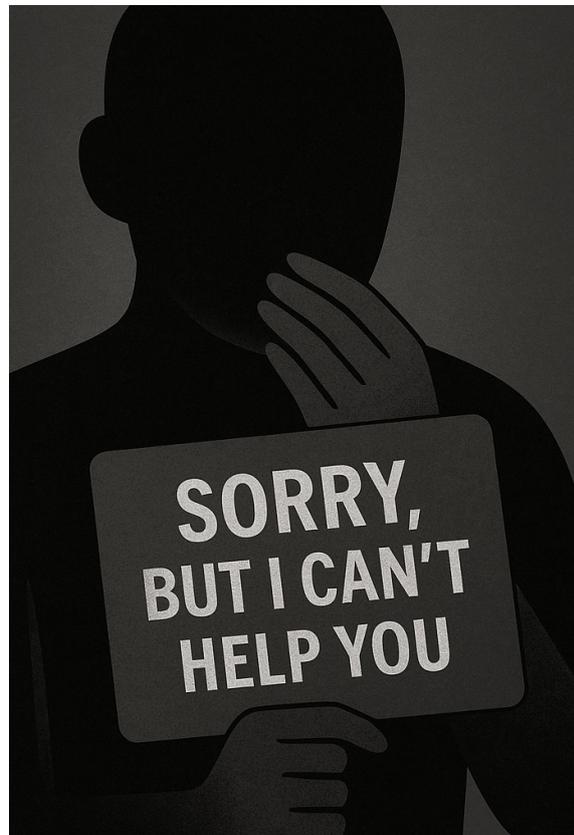
office to advocate for my colleagues, but it backfired on me. It made me get blacklisted and that's what led to my contract not being renewed.[114]

Pressure to meet production targets set by lead firms establishes an environment where even well-intentioned supervisors were deterred from supporting their staff, further isolating workers and undermining their access to care.

Isolated Good Practices

Equidem investigators also sought out good practices in supporting the wellbeing of data workers. For instance, workers employed by the BPO firm TaskUs in Colombia told Equidem investigators that they were informed upfront about the nature of their work and were provided access to on-call psychologists and guided by helpful supervisors. Dani, working as a content moderator at the BPO firm in Colombia, explained: ***They had distributed a company-wide letter saying we were going to watch explicit content and that we had resources to use in case our mental health was affected. We had a psychologist on call if we needed them The supervisor was also very helpful.[115]***

This practice of providing written information on available resources, access to on-call psychologists, and facilitating supportive supervision suggests that better practices are possible. They show that the mental health crisis in the industry is not inevitable, but the result of conscious choices by companies and demonstrate that decent work practices are not only feasible but make a meaningful difference in worker wellbeing.

ADVANCING EFFECTIVE MENTAL HEALTH SUPPORT

Current mental health provisions offered to data workers are not only inadequate, they are often part of the problem. When access to care is controlled by the same employers who enforce silence, punish emotional expression, or link counselling to performance reviews, the result is not healing – it's harm.

Superficial wellness programmes cannot solve structural violence. Trauma cannot be managed with a check-in or a company therapist, especially when workers fear retaliation for seeking help. Any serious response must treat mental health not as a "perk" or "benefit," but as a workplace right, backed by enforceable standards and free from employer control.

Principles for effective mental health support

- Mandatory, independent, trauma-informed care – not in-house therapists
- Legal protections for disclosure – including NDA exemptions for health and trauma
- Institutional safeguards against punishment for taking mental health leave
- Occupational injury recognition for PTSD, depression, and anxiety from data work
- Systemic prevention – starting with caps on exposure, work intensity, and coercive incentives

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

41

PART V
BARRIERS TO ACCOUNTABILITY

Barriers to accountability for the harms experienced by data workers – particularly those involved in content moderation and data annotation – are embedded within structural practices of secrecy, discipline, and control across the digital outsourcing industry. Denial of freedom of association, NDAs, precarious and informal employment arrangements, retaliatory dismissal, and the threat of industry-wide blacklisting together construct a regime that not only silences workers but also protects corporate interests from scrutiny.

FREEDOM OF ASSOCIATION VIOLATIONS

> *The company has not allowed workers to join or form a union at the workplace because of fears of being exposed to the world.*
>
> Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI[116]

Freedom of association is a cornerstone right because it creates the conditions for trade unions and workers' movements to intervene in structural violence by advancing, defending, and enforcing all other labour rights. Further, in the arena of legislation and policy, trade unions have the capacity to impact labour standards beyond the capacity of any individual worker. Mugo, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and Open AI, told Equidem that she would love to be part of worker organising to address exploitative wages: *I would love to be part of worker organising because we are poorly paid, especially considering the amount of work we do.*[117]

Rahi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, told Equidem that the company fired workers for trying to unionise: *We had workers who were threatened, and others sacked for trying to unionise.*[118]

In fact, in 2023, 43 workers in Kenya filed a lawsuit against Meta for unlawful redundancy. The workers said they lost their jobs with Sama, a Kenya-based firm contracted to moderate Facebook content, for organising a union. They also said they were blacklisted from applying for the same roles at another outsourcing firm, Majorel (subsequently acquired by Teleper-

formance) after Facebook switched contractors.[119]

Fatuma, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, told Equidem that when workers were terminated for trying to form a union, it reinforced a culture of fear that prevented others from unionising: *I heard that someone who tried to form a union was penalised. He was fired and sent back to his home country. Based on his story, no one after him attempted to start a union.*[120]

Chacha, a migrant worker in Kenya working as a content moderator for a BPO firm subcontracted by Meta and OpenAI, described risks associated with unionisation: *We are not allowed to form a union. We are afraid of losing our jobs, being mistreated, or even sent home.*[121]

NDAS AS INSTRUMENTS OF FORCED SILENCE

NDAs are standardised across the data annotation and content moderation sectors, often imposed as a condition of employment. While nominally intended to protect proprietary information, their application routinely extends far beyond confidentiality. In practice, Equidem found that NDAs function to suppress disclosures of:

•   Exposure to abusive and traumatic content
•   Unsafe or exploitative working conditions
•   Employer misconduct

This enforced silence has profound psychological and social consequences.

Kumao, working as a data labeller for a BPO firm in Kenya described the toll of such restrictions: *I cannot share anything with my family. They are the ones who are the closest to me, yet I cannot share anything because I have signed an NDA. I have kept everything bottled inside of me.*[122]

Research confirms that suppressing traumatic experiences significantly increases the risk of anxiety, depression, and PTSD. In this context, NDAs are not merely administrative tools but mechanisms of harm, isolating workers from both informal support networks and professional care.

42   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

Obstructing Access to Mental Health Care

NDAs obstruct access to effective mental health treatment. Workers bound by such agreements often report fearing legal repercussions if they disclose the nature of their work, even to therapists or counsellors. This severely limits their ability to access appropriate care.

Leo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, recounted: *I believe when I joined [the account]... I signed numerous contracts of non-disclosure... so I couldn't tell anyone I worked for them. Do you think I could face legal problems for telling you about the work there?*[123] Such fears are neither hypothetical nor uncommon. Workers reported withholding critical information in therapeutic settings, which renders treatment either inadequate or impossible. The result is a further entrenchment of psychological distress and isolation.

Suppressing Accountability and Reporting

Beyond its psychological impact, the NDA regime also suppresses mechanisms of accountability. Workers bound by NDAs are discouraged from reporting abuse, harassment, or unsafe working conditions – both internally and externally – due to the threat of termination, legal action, or industry-wide blacklisting.

This was evident in Equidem's fieldwork in The Philippines: all 36 data workers interviewed referenced the constraints imposed by NDAs. Several declined to answer specific questions or abruptly ended their interviews. Others who refused to participate altogether cited the risk of being identified and penalised. One worker explained that speaking out involved a calculated "cost-benefit" decision, indicating the depth of fear these agreements generate.

Intimidation and Control

NDAs serve not just to silence, but to intimidate. In a sector characterised by informal contracts and minimal protections, NDAs are often paired with threats of dismissal or legal retaliation, reinforcing existing power imbalances between employers and workers.

This climate of fear inhibits collective organising and resistance.

Adam, a migrant worker in Colombia working as a content moderator for a BPO firm in subcontracted by Meta and ByteDance, recalled: *I never thought of joining a union. I thought they were weak. We had to sign a non-disclosure agreement when we signed the contract. We couldn't talk about what happened indoors.*[124] Here, the NDA operates not only as a legal constraint but as a tool of social control, reinforcing individualisation and disempowerment within the workforce.

Obscuring Structural Harm

Finally, the widespread use of NDAs prevents broader recognition of structural harm in the industry. By silencing those most affected, NDAs hinder public awareness and inquiry into exploitative practices. They also limit the visibility of systemic abuse, protect corporate reputations, and stifle efforts toward regulatory or policy reform.

RETALIATION FOR SPEAKING OUT

> *We were threatened to be fired if we became vocal on matters affecting us as content `moderators.*
>
> Moiben, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and Open AI[125]

In a culture tightly controlled by NDAs, data workers from Colombia, Kenya, and The Philippines described high risks associated with speaking out. Data workers described being targeted by management, excluded from company forums, terminated, and blacklisted from the sector.

Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, reported that he faced repercussions from his account manager for speaking out, and that Human Resources did nothing to protect him: *I was seen as someone disruptive and not the ideal worker as I voiced my concerns and flagged behaviour and work practices that I deemed oppressive. I was personally attacked by my account manager and expressed my concerns to HR. They did nothing.*[126]

Matthew, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, said that he was

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

43

prevented from attending company town hall meetings because he asked questions about salary calculations: *I did not get the criteria used to calculate our salary. I wanted them to explain and break it down into basic and allowance. It was never received well. That was part of the reason I was blocked from entering town halls where we were supposed to raise such issues. Till I left they never explained to me how they got that figure.[127]*

Data workers from The Philippines routinely refused interviews with Equidem investigators, explaining that expressing any negative views about their employer has dire consequences. A union organiser in The Philippines explained: *Being outspoken within the sector, including advocating for one's rights, is very effectively discouraged. An outspoken person is tagged as a "complainer" and they will be prioritised for dismissal. The dismissal can be done creatively, including by intentional down-grading or demotion, effectively forcing the employee to resign or be sacked due to "low performance" which can be subjective. I know of one case where an employee was asked to go to a meeting room with their team leader where they were given a pen and paper and forced to write a resignation letter on the spot.[128]*

Termination of workers who raise concerns not only silences workers with the courage to speak out, but also enforces a culture of silence, driven by fear of retaliation. Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, explained: *I have seen others who tried to voice work-related concerns get fired. So, I have resorted to being silent.[129]*

INEFFECTIVE GRIEVANCE REDRESSAL CHANNELS

Data workers employed by BPO firms subcontracted by Meta in Colombia, Ghana, and Kenya all described ineffective grievance and dispute resolution channels.

Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, explained: *There are no clear dispute resolution mechanisms with management. The only available space is talking to the supervisor, which has not been effective.[130]*

Dumba, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, also reported that there were no grievance channels: *There is no specific body to handle grievances. We can talk to our team leaders, if needed.[131]*

Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, reported that helpdesks at the office fail to address workers' concerns: *They have a helpdesk email and representatives for groups like Women, LGBT+, and Mothers, but they don't work well. They rarely respond to emails, and when they do, they twist words to avoid solving issues, claiming there isn't a problem.[132]*

Lead firm failure to establish and implement meaningful corporate human and labour rights commitments

Meta and OpenAI have established formal policies outlining their commitments to human and labour rights. This report, however, documents rampant violations of the Meta and OpenAI policies, confirming the complete inefficacy of human rights standards that make hollow commitments without binding channels for enforcement. Notably, ByteDance and Remotasks lack publicly available information detailing their commitments in these areas.

Meta

In March 2021, Meta introduced its inaugural corporate human rights policy, aligning with the UNGPs. This policy outlines Meta's commitment to respecting human rights across its operations and includes provisions for annual public reporting on human rights issues.

Open AI

OpenAI has articulated its commitment to ethical business practices through its Supplier Code of Conduct, updated in October 2023. This document mandates that suppliers comply with all applicable laws and uphold high standards in areas such as labour rights, health and safety, and ethical conduct.

44   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

In October 2024, OpenAI published its Raising Concerns Policy, designed to protect employees' rights to report issues related to AI safety, legal compliance, or company policies. The policy offers multiple channels for raising concerns and explicitly prohibits retaliation against employees who do so. It also affirms employees' rights to make disclosures to government agencies, including the U.S. National Labor Relations Board and the U.S. Equal Employment Opportunity Commission.

THE UN GUIDING PRINCIPLES ON BUSINESS AND HUMAN RIGHTS

Under international standards on business and human rights, all companies must respect all human rights. This includes labour rights. This is articulated in the UN Guiding Principles on Business and Human Rights (UNGPs), an internationally accepted set of standards endorsed by the UN Human Rights Council. The responsibility to respect human rights extends not only to the company's own activities but also to its business relationships, such as relationships with sub-contractors. The UNGPs note that a company's "activities" include both actions and omissions.[133]

To meet their responsibility to respect human rights, companies should have in place a human rights due diligence process to identify, prevent, mitigate and – where necessary – redress human rights abuses connected to their operations. Multinational technology companies are responsible not only for how they treat the workers they employ directly, but also to act with due diligence to ensure that companies they sub-contract from not abuse the rights of people working for them.

Table 17: Meta and OpenAI violations of internal human and labour rights standards

	Meta violations of the UN Guiding Principles on Business and Human Rights/Meta corporate human rights policy	OpenAI Supplier Code of Conduct Violations
Target pressure and industrial discipline practices to drive accelerated work		
Unreasonable targets leading to high stress and extended working hours	Principles 12–13, 17	Working Hours, Health and Safety, Humane Treatment
Threats of termination for failure to meet targets	Principles 13, 17, 23(b)	-
Extended hours and overwork	Principles 12–13, 17	Working Hours, Health and Safety

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

45

Unpaid overtime	Principles 13, 17, 20	Wages and Benefits, Working Hours
Physical harms		
Physical impacts of exposure to violent content as a feature of work processes	Principles 13, 17, 18	Health and Safety
Extended hours and overwork	Principles 12–13, 17	Working Hours, Health and Safety
Denial of leave	Principles 12–13, 23(a)	-
Denial of sick leave	Principles 12, 13, 23(a)	-
Psychological harms		
Ongoing exposure to violent and sexually explicit content as a feature of work processes	Principles 13, 17, 18	Health and Safety, Humane Treatment
Extended working hours conducting isolated repetitive tasks as a feature of work processes	Principles 13, 17, 18	Health and Safety, Humane Treatment
Stress and anxiety from decision fatigue as a feature of work processes	Principles 13, 17, 18	Health and Safety, Humane Treatment
Verbal abuse and workplace bullying	Principles 12, 13, 17, 23(b)	-
Sexual harms		
Exposure to pornographic and violent sexual content resulting in sexual trauma as a feature of work processes	Principles 13, 17, 18	Humane Treatment, Health and Safety, Non-Discrimination
Sexual harassment	Principles 13, 17, 23(a)(b)	-
Inadequate psychological support		
Failure to provide occupational health services, including psychological support	Principles 12–13, 18, 23(a)	Health and Safety
Economic harms		
Unpaid overtime	Principles 13, 17, 20, 23(a)	Wages and Benefits, Working Hours
Task-based low wages	Principles 13, 17, 20	Wages and Benefits
Firm relocation without compensation	Principles 13, 17, 20, 22	-
Freedom of association violations		
Retaliation for organising a union	Principles 12, 13, 18, 23(a)(b)	Freedom of Association, Non-Retaliation, Ethics

CONCLUSION AND RECOMMENDATIONS

> *I was never under 55 kg after I was 16 years old – but at 22, I got to 45 kg working for them. I was losing hair; my nails didn't grow anymore. There was a month that I just did all the drugs I could to not think about the things I was watching, and also what was happening to me.*
>
> Sofia, a migrant worker in Colombia, working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance[134]

This report exposes a global digital labour system built on institutionalised harm. Far from the sleek branding of AI as clean, contactless, and futuristic, the reality is one of deeply human labour – exhausting, precarious, and often traumatising. Behind every AI product and moderated post is a data worker facing the brunt of toxic content, algorithmic pressure, and corporate opacity.

Workers across Colombia, Kenya, and The Philippines have made one thing clear: the violence is not only in the content they are forced to review, but in the labour system itself – one that inflicts psychological trauma, denies basic protections, retaliates against organising, and leaves them legally and economically exposed.

Despite their essential contributions to the digital economy, these workers remain invisible in supply chains governed by profit, secrecy, and denial – denial of harm, of responsibility, and of workers' basic rights. They are treated as expendable tools in the race to train, test, and deploy AI systems at scale. Yet their voices – documented throughout this report – are unambiguous. The harm is real. The cost is human.

Addressing this crisis requires more than technical tweaks or marginal reforms. It demands a structural shift in how digital labour is governed, who bears responsibility, and what protections are guaranteed.

We are at a critical moment. As AI becomes more central to society, so too does the labour that sustains it. The choice ahead is stark: entrench a digital economy that profits from invisible suffering – or build one grounded in safety, dignity, and justice for those who power it. This report makes clear which path we must take – and what must happen next to get there.

RECOMMENDATIONS FOR REDUCING HARM AND REFORMING THE DIGITAL LABOUR SUPPLY CHAIN

This section presents a structured roadmap to improve working conditions for data workers in the AI and content moderation supply chains. It breaks down what needs to be done now, in the near future, and over the long term, with each recommendation tailored to the specific responsibilities of key actors.

We organise these recommendations along two dimensions:

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

47
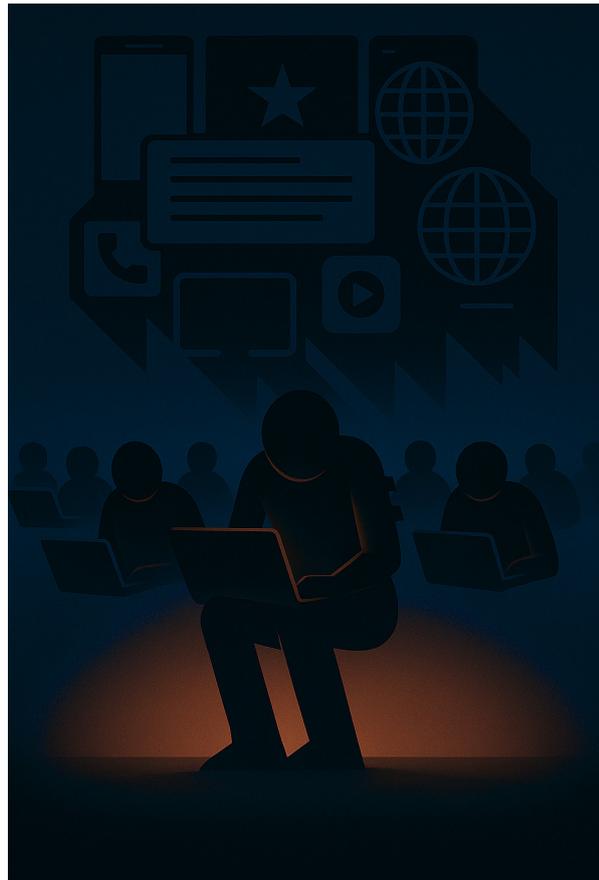
URGENCY - HOW QUICKLY ACTION IS NEEDED

- **Immediate measures (0–6 months):** Practical steps that can be implemented right away to reduce harm and prevent further abuse.
- **Short- to medium-term reforms (6–18 months):** Legal, policy, and institutional changes that address deeper systemic failures.
- **Long-term structural change (18+ months):** Transformational shifts to redesign the digital labour system and confront its root inequalities.

RESPONSIBILITY - WHO MUST ACT

A. LEAD FIRMS (LIKE META, OPENAI, BYTEDANCE)

B. INVESTORS AND FINANCIAL BACKERS OF AI AND TECH

C. THE INTERNATIONAL LABOUR ORGANIZATION

D. HEADQUARTER ECONOMIES (COUNTRIES WHERE LEAD FIRMS ARE BASED)

E. STATES WITH DIGITAL LABOUR WORKFORCES (LIKE COLOMBIA, KENYA, THE PHILIPPINES)

F. AI RESEARCHERS AND ACADEMIC INSTITUTIONS

I. IMMEDIATE AND ACTIONABLE MEASURES (0–6 MONTHS)

These measures are designed to address the most pressing harms faced by digital labour workers and can be implemented immediately. They focus on reducing psychological trauma, preventing further rights violations, and laying the groundwork for more substantial reforms.

For Lead Firms

1. **Uphold Freedom of Association and Collective Bargaining Rights:**
Require all subcontractors and third-party firms to respect workers' rights to organise, join unions, and engage in collective bargaining. Retaliation for union activity must be explicitly banned and monitored.

2. **Prohibit Team-Based Incentive and Punishment Systems:**
Abolish team-based performance bonuses and disciplinary frameworks that penalise all members for individual underperformance. These systems foster hostility, enable workplace bullying, and create additional stress in already high-pressure environments.

3. **Ensure Access to External, Trauma-Informed Mental Health Care:**
Provide workers with free and confidential access to licensed mental health professionals trained in treating trauma from graphic and sexually explicit content exposure. Mental health support must be offered independently of the employer to ensure trust and efficacy.

4. **Suspend the Use of Gag Clauses in Employment Contracts:**
Immediately suspend the use of NDAs that prevent workers from discussing job-related trauma, workplace conditions, or rights violations. Workers must retain the freedom to report harm, seek care, and organise without legal intimidation.

5. **Mandate Written Contracts and Paid Training for All Workers:**
Ensure all subcontracted or outsourced workers are employed under formal, written contracts that clearly outline job responsibilities, protections, and compensation. Eliminate unpaid training periods, which constitute unrecognised labour and exploit vulnerable workers.

For Headquarter Economies of Lead Firms

1. **Enforce Mandatory Transparency in Global Supply Chains:**
Require all tech companies headquartered within the jurisdiction to disclose the full list of BPO firms, labour agencies, and downstream digital labour contractors engaged in content moderation and data annotation.

For AI Researches in Academia and Industry

1. **Establish Labour Ethics Review for AI Projects:**
Develop and institutionalise protocols – akin to Institutional Review Boards (IRBs) in human subjects research – that assess the labour conditions under which training data is collected, labelled, and moderated.

2. **Engage Directly with Workers and Labor Scholars:**
Foster collaborative research models that include content moderators, data annotators, and labour experts in shaping ethical AI development practices. Ensure findings are informed by frontline realities.

3. **Refuse to Use Exploitatively Labelled Datasets:**
Implement a policy of ethical refusal in research and product development. Datasets produced under conditions of coercion, overwork, or trauma must be excluded from training AI systems to ensure that ethical principles are upheld across the lifecycle of machine learning.

For Investors

1. **Require Mandatory Human Rights Due Diligence (HRDD):**
Condition all future investments in AI, platform companies, and digital services on demonstrated compliance with international labour standards. HRDD must include assessments of subcontractor practices, mental health safeguards, and freedom of association protections.

ADVANCING EFFECTIVE MENTAL HEALTH SUPPORT

Current mental health provisions offered to data workers are not only inadequate, they are often part of the problem. When access to care is controlled by the same employers who enforce silence, punish emotional expression, or link counselling to performance reviews, the result is not healing – it's harm.

Superficial wellness programmes cannot solve structural violence. Trauma cannot be managed with a check-in or a company therapist, especially when workers fear retaliation for seeking help. Any serious response must treat mental health not as a "perk" or "benefit," but as a workplace right, backed by enforceable standards and free from employer control.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

49

> **Principles for effective mental health support**
>
> - Mandatory, independent, trauma-informed care – not in-house therapists
> - Legal protections for disclosure – including NDA exemptions for health and trauma
> - Institutional safeguards against punishment for taking mental health leave
> - Occupational injury recognition for PTSD, depression, and anxiety from data work
> - Systemic prevention – starting with caps on exposure, work intensity, and coercive incentives

II. SHORT- TO MEDIUM-TERM REFORMS (6–18 MONTHS)

These reforms address the systemic regulatory and institutional failures that have enabled exploitative conditions in the digital labour supply chain. Over the next 6 to 18 months, actors across sectors must act to embed stronger protections, accountability mechanisms, and fairer standards into the global digital economy.

FOR LEAD FIRMS

1. **Adopt long-term stable contracting relationships with BPO providers to establish conditions to set and enforce decent work standards:**
   Tasks involving prolonged exposure to violent, sexually explicit, or abusive content – such as inappropriate interactions with children or gender-based violence content – must be undertaken in context of long-term stable contracting relationships with BPO providers in order to ensure they come with full employment protections, mental health coverage, and long-term support structures.

2. **Set specific guidelines on Bonus and Compensation Structures that meet decent work standards and ensure adequate pricing structures for enforcement:**
   Standards should revise remuneration systems across business operations to ensure stability and fairness. Performance bonuses must be capped at no more than 25% of total compensation to avoid creating perverse incentives, coercive overwork, and psychological pressure. Compensation should prioritise well-being over speed.

3. **Establish independent grievance mechanisms:**
   Develop and fund independent, third-party channels through which workers can report abuse, harassment, or unsafe conditions without fear of retaliation. These mechanisms should be accessible in multiple languages, culturally sensitive, and protected from employer interference.

FOR INVESTORS

1. **Demand Transparency and Legal Compliance from Investees:**
   Require portfolio companies to provide detailed reports on subcontractor hiring and employment practices, employment models, and adherence to ILO standards. Ongoing investment should be contingent on demonstrated improvement and compliance.

FOR THE INTERNATIONAL LABOUR CONFERENCE (ILC)

1. **Recognise Content Moderation and Data Annotation as High-Risk, Essential Work:**
   Issue a formal classification that acknowledges the psychological and occupational hazards faced by workers in digital content review and AI data annotation. This recognition will guide legal protections and standard-setting processes across member states.

FOR HEADQUARTER ECONOMIES

1. **Enforce Ethical Procurement Policies in the Public Sector:**
   Bar government contracts from being awarded to companies that rely on abusive outsourcing practices. Require bidders to demonstrate compliance with international labour laws and supply chain transparency obligations.

2. **Advance Joint Employer Liability Legislation at National and Regional levels:**
   Enact legal frameworks that hold lead firms jointly responsible for labour rights violations perpetrated by their subcontractors, regardless of formal employment structure. Adopt a strong enforceable EU Corporate Sustainability Due Diligence Directive (CSDDD) that ensures that corporations cannot evade accountability through outsourcing.

50   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

III. LONG-TERM STRUCTURAL CHANGES (18+ MONTHS)

The following recommendations reflect the need for deep, systemic transformation in the architecture of digital labour. These changes aim not only to reduce harm, but to rebalance power across the AI supply chain, ensuring that workers are recognised as essential contributors to technological development and protected by the same standards afforded to other sectors. Structural change requires rethinking business models, regulatory frameworks, and the ethical foundations of AI development itself.

FOR LEAD FIRMS

1. **Redesign Platforms to Reduce Human Exposure to Harmful Content:**
   Shift from reactive human moderation to proactive platform design. This includes building in upstream filters, automated content detection with human-in-the-loop ethical checks, and architectural redesigns that limit the circulation of toxic material. The long-term goal is to minimise reliance on human exposure to trauma while maintaining accountability and oversight.

2. **Establish a Global Recovery and Reparations Fund for Affected Workers:**
   Lead firms must finance a trauma recovery fund for current and former workers who suffer long-term psychological or physical harm from moderation and annotation work. This fund should be independently managed and provide access to long-term mental health care, vocational retraining, and income support.

FOR THE INTERNATIONAL LABOUR CONFERENCE (ILC)

1. **Recognise content moderators and data labellers** as digital platform economy workers and advance a binding convention that extends ILO labour standards protection to this category of work.

FOR AI RESEARCHERS IN ACADEMIA AND INDUSTRY

1. **Integrate Labour Rights into AI and Data Science Education:**
   Revise computer science, machine learning, and AI curricula to include modules on labour ethics, digital worker rights, and the political economy of AI. Training future technologists in these areas is essential to building a more equitable and sustainable AI ecosystem.

2. **Build Public, Non-Extractive Data Labelling Platforms:**
   Establish platforms – supported by public funding or nonprofit partnerships – where communities can voluntarily contribute to AI training in ways that are transparent, collectively governed, and non-exploitative. These platforms should offer alternatives to commercial models that rely on precarious labour.

FOR INVESTORS

1. **Divest from Non-compliant or Repeat-Offender Companies:**
   Develop and enforce investment screening mechanisms that identify and exclude firms with consistent patterns of labour rights violations, unsafe working conditions, or systemic non-compliance with human rights standards.

2. **Adopt and Promote "Well-being First" Investment Strategies:**
   Shift portfolio strategy to prioritise worker well-being as a core investment criterion. This includes supporting companies that centre health, safety, and fair labour practices in their AI and digital operations, as well as incentivising ethical innovation in supply chain practices.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

51

ANNEXURE - COMPANY RESPONSES

META

RESPONSE RECEIVED ON 29 APRIL 2025

Dear Mr. Qadri,

Many thanks for your letter of April 17th on Re: Occupational Health and Safety Risks and Workplace Violence at Meta.

We appreciate the opportunity to respond to the concerns you have raised in the letter ahead of your forthcoming report.

We take the support of content reviewers seriously which is why we have clear contracts with each of the companies that help review content on Facebook and Instagram that detail our expectations in a number of areas, including counselling, training and other support.

We require all of the companies we work with to provide 24/7 on-site support with trained practitioners, an on-call service, and access to private healthcare from the first day of employment. They are also contractually obliged to pay their employees who review content on Facebook and Instagram above the industry standard in the markets they operate.

We respect the right of our vendor's employees to organize and do not oppose or inhibit their right to unionize.

I hope that this information is useful.

Warmly,
Iain
Iain Levine
Human rights policy team,
Meta

52

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

OPEN AI

RESPONSE RECEIVED ON 5 MAY 2025

Dear Mustafa,

Thank you for reaching out and for taking the time to share your detailed letter (attached) outlining the serious concerns raised by Equidem regarding worker wellbeing within our vendor ecosystem. We want to assure you that we take these matters extremely seriously.

At OpenAI, we are committed to upholding human rights and fostering safe, respectful working conditions—including for those supporting our work via third-party partners. The findings you've presented are deeply concerning, and we are treating them with the urgency and care they deserve.

To better understand the context and ensure we can thoroughly investigate, I would welcome the opportunity to connect directly with you and your team. I'd appreciate the chance to schedule a conversation at your earliest convenience to learn more about your findings, clarify any underlying facts, and ensure we are actively engaging on the appropriate next steps.

Please let us know a few times that work for you and we'll coordinate accordingly. We appreciate your partnership in supporting transparency and accountability, and we look forward to speaking soon.

Thanks,

Brett Klinker
Extended Workforce Team
go/extendedworkforce

Sent: 15 May 2025 22:20
To: Mustafa Qadri <mqadri@equidem.org>
Cc: Equidem General Enquiries <info@equidem.org>
Subject: Re: Subject: Re: Worker Safety and Wellbeing Concerns

Hi Mustafa,

Thank you for taking the time to meet with me last week. As I mentioned in our discussion, we appreciate you sharing the findings of your investigation. We respect your work and take these allegations extremely seriously.

At OpenAI, we are committed to creating a safe and positive working environment for everyone involved in building our technology, including those employed by our suppliers.

As such, I wanted to share our supplier code of conduct, which sets out clear expectations for the ethical and fair treatment of all workers, whether directly employed or part of an outsourced team. This code covers critical areas like workplace safety, fair compensation, non-discrimination, and respect for workers' rights.

Additionally, since May 2024, our major outsource suppliers have signed up to an additional worker conditions addendum, which reinforces our expectations of partners to be aligned with our high standards around worker conditions, pay and support. We are in the process of rolling this out more broadly across our supplier network.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

53

After our call, we carefully reviewed the limited information you were able to provide. Unfortunately, as I flagged may be the case, the details shared were not sufficient for us to fully investigate the claims. Despite this, we conducted an investigation based on the available information and following inquiries with our suppliers, we found no evidence to support the claims.

Lastly, as discussed during our call, I'd like to reiterate our request for additional information to help us conduct a full and thorough investigation. If you're able to share any additional information, I can assure you, we will investigate this further. We continue to uphold the highest standards across our operations and with our suppliers, and are grateful to anyone raising allegations of concerning behavior so we can investigate it and take appropriate action.

We remain committed to OpenAI's goal that anyone working within our ecosystem does so in a safe, supportive, and respectful environment.

Thank you again for sharing your findings and please don't hesitate to contact me if you have further information.

BYTEDANCE

No response as of May 22, 2025

REMOTASKS

RESPONSE FROM JACKIE FROM SCALE AI ON APRIL 21, 2025

Hi - Thank you for sending this over. Do you have a few minutes to discuss this on the phone today or tomorrow? Time zone wise, I am in California on PT.

I am interesting in better understanding the end report. Will this be just focused on Scale AI, or other companies as well?

RESPONSE FROM JACKIE ZUPSIC ON 24 APRIL 2025

Mustafa and Shikha,

Thank you for taking the time to meet with us yesterday. Please find a full statement below to accompany your report publication. Your questions are addressed within the statement itself.

Please let me know if you have any questions. We would be grateful to know when you intend to publish the report as well - once that gets settled.

Best,
Jackie

54   Scroll. Click. Suffer
      The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

Statement & background

Understanding Remotasks: Remotasks pays contributors per task completed, offering flexible work options to people around the world. When contributors complete tasks, they're training our customers' AI systems. These customers include leading organizations across various industries that rely on high-quality data to improve their AI models. We prioritize offering flexible opportunities that allow contributors to earn supplemental income on their own schedule – whether in the evenings, on weekends, or part-time – while playing a critical role in the advancement of real-world AI applications.

Fair wages and Labor Protections: Pay rates are transparently listed upfront and guardrails are put in place to ensure that workers' pay rates meet or exceed minimum wage in all geographies we operate. We also provide workers with an estimate of how long each task is expected to take. If a contributor notices a payment issue, they can contact our support team to review and make any necessary pay adjustments. In Q1 2025, approximately 75% of support tickets, across all categories, were resolved within 48 hours. In regions where we work through third parties, we monitor compliance with labor standards and contractually require that all partners follow strict codes of conduct.

Developing Safer AI systems: One area where our customers engage our Remotasks services involves supporting the development of safer AI systems. Training models to prevent harmful and abusive content is essential for the safe development of AI. While some AI safety projects involve sensitive content, we always give contributors advance notice and the option to opt out at any time. To support them in this important work, Remotasks has implemented a range of policies and safeguards to help prevent psychological harm, especially when projects involve sensitive or explicit content. Contributors are given advance notice about the nature of their tasks and have the option to opt out at any time. Opting out does not affect future availability of work. To support their well-being, we offer access to health and wellness programs, including weekly wellness sessions and dedicated mental health resources. Additional safeguards include a 24/7 anonymous hotline and a ticketing system for reporting concerns. Ultimately, many people choose to take on this work because they believe in the mission and importance of training AI for the greater good.

Support and Appeals: All contributors have access to numerous support options. These include a 24/7 anonymous hotline, a ticketing system available directly through the platform, and live support chat for urgent needs. All complaints are reviewed by dedicated support staff and escalated when appropriate to ensure a thorough and fair resolution process. Any claims of unfair treatment are taken seriously and investigated.

Ongoing Improvements: Remotasks takes worker feedback seriously and invests heavily into ongoing improvements to the platform. Recent changes include:

- A new pay structure that rewards contributor work, as well as a new pay dashboard that gives additional context to how earnings were calculated each period.

- A consolidated platform to make it easier for contributors to ask questions, find support, and communicate about their work.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

55

- A platform update that improves access to tasks, creating a more transparent experience, and supporting fairer task distribution. This includes automatic secondary project offerings to minimize idle time, automated timers that return inactive work to the queue, and the introduction of base payouts to account for pre-annotation effort.

- A new Account Support function with clearer deactivation guidelines and a simplified appeals process, and increased transparency around work allocation.

- Updated Terms to provide a 30-day advance notice of contract changes and clarified that task participation is voluntary and does not impact contributor ratings.

- Updated Community Guidelines to include a zero-tolerance policy for harassment, affirm the voluntary nature of task participation, and provide more accessible pathways for account deactivation support.

Jackie Zupsic She/Her
Executive Vice President
tuskstrategies.com **|** (805) 698-5060

56   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

REFERENCES

1. Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

2. Kepios, Global Social Media Statistics, Data Reportal, accessed March 28, 2025, https://datareportal.com/social-media-users.

3. Meghan McCarty Carino and Daniel Shin (2023). "The human labor behind AI chatbots and other smart tools," MarketPlace Tech, March 21, 2023.

4. Psoriasis is a skin condition that causes flaky patches of skin which form scales. "Overview: Psoriasis", National Health Service (NHS) United Kingdom website, accessed April 9th, 2025: https://www.nhs.uk/conditions/psoriasis/.

5. Pamela, a Filipina national working as a content moderator for a BPO firm subcontracted by a social media platform in The Philippines, interviewed in August 2024.

6. Camilia, working as a content analyst for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 24, 2024.

7. Mile, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

8. See ILO Occupational Safety and Health Convention, 1981 (No.155), Article 3(e): "the term health, in relation to work, indicates not merely the absence of disease or infirmity; it also includes the physical and mental elements affecting health which are directly related to safety and hygiene at work."

9. Evans, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 14, 2024.

10. Aziz, working as a quality analyst for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 10, 2024.

11. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

12. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on September 4, 2024.

13. Alamini, working for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on 3 September 2024.

14. A.J. Bridges, R Wosnitzer, E. Scharrer, C. Sun, and R. Liberman (2010). "Aggression and sexual behavior in best-selling pornography videos: A content ana lysis update." Journal of Sex Research, 47(4), 327-339.

15. Shelly Grabe, Janet Shibley Hyde, and Sara M. Lindberg (2007). "Body objectification and depression in adolescents: The role of gender, shame, and rumination." Psychology of Women Quarterly 31.2 (2007): 164-175.

16. Mike Allen, Dave D'alessio, and Keri Brezgel (1995). "A meta analysis summarizing the effects of pornography II aggression after exposure." Human communication research 22.2 (1995): 258-283.

17. Iman, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 11, 2024.

18. Miguel, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 28, 2024.

19. Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, interviewed on September 11, 2024.

20. Fanaka, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 3, 2024.

21. Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 19, 2024.

22. Rahi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 14, 2024.

23. Reuters, Content moderators sue Meta over alleged 'union busing' in Kenya, March 21, 2023, https://www.reuters.com/technology/content-moderators-sue-meta-over-alleged-union-busting-kenya-2023-03-20/.

24. Onyango, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 14, 2024.

25. Chapman BP, Fiscella K, Kawachi I, Duberstein P, Muennig P. Emotion suppression and mortality risk over a

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

57

12-year follow-up. J Psychosom Res. 2013 Oct;75(4):381-5. doi: 10.1016/j.jpsychores.2013.07.014. Epub 2013 Aug 6. PMID: 24119947; PMCID: PMC3939772.

26. Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 19, 2024.

27. Sana Ahmad and Martin Krzywdzinski (2022). "Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains," In: Mark Graham and Fabian Ferrari (Ed.), Digital Work in the Planetary Market, The MIT Press, International Development Research Centre, Cambridge, MA, Ottawa, pp. 77-95.

28. Dev Nathan, Shikha Silliman Bhattacharjee, Purushottam Kumar, Sukhpal Singh, Padmini Swaminathan, Rahul S., Reversed Subsidies: Gender, Labour, and Environmental Injustice in Garment Value Chains, Cambridge University Press, 2022.

29. ProColombia. (2023). "Outsourcing of Services (BPO). Invest in Colombia."

30. Conectys (2022). "Colombia vs Top LATAM Markets: Unveiling the BPO Powerhouse Winning Potential."

31. EN-Contact (2022). "From Accenture to Teleperformance: Who Are the Whizzes of Content Moderation?"

32. Justin Hendrix, "Checking on the Progress of Content Moderators in Africa," TechPolicy.Press, December 3, 2023.

33. Jacqueline Wambui Wamai, Maureen Chadi Kalume, Monicah Gachuki and Agnes Mukami, "A new social contract for the social media platforms: prioritizing rights and working conditions for content creators and moderators," International journal of labour research, Vol. 12 (1-2), pp. 98-108.

34. A. Drootin (2021). "Community Guidelines": The Legal Implications of Workplace Conditions for Internet Content Moderators." Fordham Law Review, 90(3), 831.

35. Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

36. See Linkedin profile for Scale AI, accessed April 4, 2025, https://www.linkedin.com/company/scaleai/

37. Elinah, working as a data labelling trainer for Remotasks in Kenya, interviewed on September 3, 2024.

38. Dhania, working as a data labelling trainer for Remotasks in Kenya, interviewed on September 11, 2024.

39. Linet, working as a data labelling trainer for Remotasks in Kenya, interviewed on September 4, 2024.

40. Elinah, working as a data labelling trainer for Remotasks in Kenya, interviewed on September 3, 2024.

41. See Asia Floor Wage Alliance (2024), Threaded Insecurity: The Spectrum of Informality in Garment Supply Chains, written by Shikha Silliman Bhattacharjee.

42. Thomas Stackpole, "Content Moderation is Terrible by Design," Harvard Business Review, November 9, 2022.

43. Bitok, working as a content moderator for a BPO firm subcontracted by ByteDance, interviewed on September 10, 2024.

44. Pablo, a migrant worker in Colombia working as a content moderator for BPO firm subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

45. Mile, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

46. Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, interviewed on September 11, 2024.

47. Equidem interviews with Filipina workers in August 2024. Memo on file with authors

48. Edna, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

49. Thomas Stackpole, "Content Moderation is Terrible by Design," Harvard Business Review, November 9, 2022.

50. Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 27, 2025.

51. Evans, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 14, 2024.

52. Aziz, working as a quality analyst for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 10, 2024.

53. Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, interviewed on September 11, 2024.

58 Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

54. Dumba, a migrant worker in Ghana working as a social media content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 6, 2024.

55. Kojo, working as a content analyst in Ghana for a BPO firm subcontracted by Meta and ByteDance, interviewed on June 6, 2024.

56. Evans, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 14, 2024.

57. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

58. Alarico, migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

59. Fanaka, working as a content moderator for a BPO firm subcontracted by Meta and OpenAI, interviewed on September 3, 2024.

60. Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on 27 August 2024.

61. Antonio, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

62. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

63. Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, interviewed on September 11, 2024.

64. Edna, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

65. Moiben, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and Open AI, interviewed on September 3, 2024.

66. Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

67. Daina, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and Open AI, interviewed on August 15, 2024.

68. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

69. Kijani, working as a quality analyst for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 11, 2024.

70. Moiben, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and Open AI, interviewed on September 3, 2024.

71. Chacha, a migrant worker in Kenya working as a content moderator for a BPO firm subcontracted by Meta and OpenAI, interviewed on September 2, 2024.

72. Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 11, 2024.

73. Evelyne Musambi, "Kenya-based content moderators mourn Nigerian colleague who 'was desperate to go home'," AP News, March 12, 2025.

74. Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

75. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 4, 2024.

76. Pablo, a migrant worker in Colombia working as a content moderator form a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 27, 2024

77. Dumba, a migrant worker in Ghana working as a social media content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 6, 2024.

78. Aziz, working as a quality analyst for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 10, 2024.

79. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 4, 2024.

80. Samuel, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 29, 2024.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

59

81. Camilia, working as a content analyst for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 24, 2024.

82. Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed in August 2024.

83. Fanaka, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 3, 2024.

84. Iman, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 11, 2024.

85. Daudi, working as an AI worker for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 10, 2024.

86. Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 11, 2024.

87. Miguel, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 28, 2024.

88. Evans, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 14, 2024.

89. Sara Berg (2021), "What doctors wish patients knew about decision fatigue." American Medical Association, November 19, 2021.

90. Alarico, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed in August 2024.

91. Evans, a migrant worker in Ghana working for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 14, 2024.

92. Onyango, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 3, 2024.

93. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 4, 2024.

94. A.J. Bridges, R Wosnitzer, E. Scharrer, C. Sun, and R. Liberman (2010). "Aggression and sexual behavior in best-selling pornography videos: A content ana lysis update." Journal of Sex Research, 47(4), 327-339; Niki Fritz et al. (2020). "A descriptive analysis of the types, targets, and relative frequency of aggression in mainstream pornography." Archives of Sexual Behavior (2020): 1-13.

95. Shelly Grabe, Janet Shibley Hyde, and Sara M. Lindberg (2007). "Body objectification and depression in adolescents: The role of gender, shame, and rumination." Psychology of Women Quarterly 31.2 (2007): 164-175.

96. Mike Allen, Dave D'alessio, and Keri Brezgel (1995). "A meta analysis summarizing the effects of pornography II aggression after exposure." Human communication research 22.2 (1995): 258-283.

97. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 4, 2024.

98. Edna, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, interviewed on September 03, 2024.

99. Iman, working for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 11, 2024.

100. Aziz, working as a quality analyst at a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 10, 2024.

101. Kijani, working as a quality analyst for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 11, 2024.

102. Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, interviewed on September 11, 2024.

103. Iman, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 11, 2024.

104. Evans, a migrant worker in Ghana working for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 14, 2024.

105. Equidem interviews with Filipino workers in August 2024. Memo on file with authors.

106. Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

107. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and

60   Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

equidem.org

ByteDance, interviewed on September 4, 2024.

108. Philip, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on September 5, 2024.

109. Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

110. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

111. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

112. Samuel, working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 29, 2024.

113. Pablo, a migrant worker in Colombia working as a content moderator for a BPO firm in Colombia subcontracted by Meta and ByteDance, interviewed on August 27, 2024.

114. Hamidi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and ByteDance, interviewed on September 11, 2024.

115. Dani, working as a content moderator at a BPO firm in Colombia, interviewed on August 26, 2024.

116. Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 11, 2024.

117. Mugo, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 4, 2024.

118. Rahi, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 14, 2024.

119. Reuters, Content moderators sue Meta over alleged 'union busing' in Kenya, March 21, 2023, https://www.reuters.com/technology/content-moderators-sue-meta-over-alleged-union-busting-kenya-2023-03-20/.

120. Fatuma, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 5, 2024.

121. Chacha, a migrant worker in Kenya working as a content moderator for a BPO firm subcontracted by Meta and OpenAI, interviewed on September 2, 2024.

122. Kumao, working as a data labeller for a BPO firm in Kenya, interviewed on September 14, 2024.

123. Leo, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 1, 2024.

124. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

125. Moiben, working as a content moderator for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 3, 2024.

126. Adam, a migrant worker in Colombia working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 3, 2024.

127. Matthew, working as a data labeller for a BPO firm in Kenya that is subcontracted by Meta and OpenAI, interviewed on August 26, 2024.

128. Equidem interviews with Filipina workers in August 2024. Memo on file with authors.

129. Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 19, 2024.

130. Haki, working as a data labeller for a BPO firm in Kenya subcontracted by Meta and OpenAI, interviewed on September 19, 2024.

131. Dumba, a migrant worker in Ghana working as a content moderator for a BPO firm subcontracted by Meta and ByteDance, interviewed on August 6, 2024.

132. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 4, 2024.

133. For more information on the applicability of the UNGPs to technology-based businesses see "The UN Guiding Principles in the Age of Technology: a B-Tech Foundational Paper", UN Human Rights, accessed on April 9th, 2025:
https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/introduction-ungp-age-technology.pdf.

134. Sofia, a migrant worker in Colombia working as a quality analyst for a BPO firm subcontracted by Meta and ByteDance, interviewed on September 4, 2024.

equidem.org

Scroll. Click. Suffer
The Hidden Human Cost of Content Moderation & Data Labelling

61

equidem