

THE PEOPLE BEHIND THE SCREENS:

WHY TECH COMPANIES NEED NEW PROTOCOLS
FOR SAFE CONTENT MODERATION



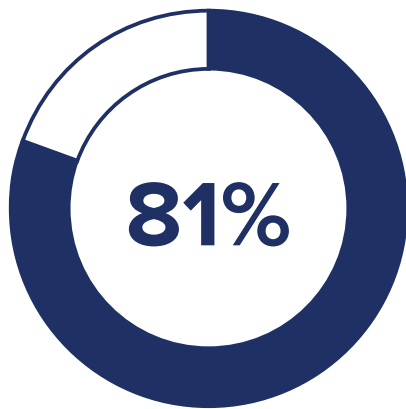
UAI
global
union

CONTENT

SECTION 1 – EXECUTIVE SUMMARY	3
SECTION 2 – CONTENT MODERATION THE HIDDEN WORKFORCE KEEPING THE INTERNET SAFE	5
<i>The business of content moderation: Outsourcing accountability</i>	6
SECTION 3 – FOUR KEY CHALLENGES FACED BY CONTENT MODERATORS	7
1. <i>Egregious content and vicarious trauma</i>	7
2. <i>Managed by machines: The relentless pace of content moderation</i>	9
3. <i>Lack of training and support</i>	9
4. <i>Barriers to unionization</i>	10
<i>Case study: How other trauma-exposed professions protect workers</i>	11
<i>Key Lessons for the Content Moderation Industry:</i>	12
SECTION 4 – EIGHT PROTOCOLS FOR SUSTAINABLE CONTENT MODERATION	13
CONCLUSION – A CALL TO ACTION	16

SECTION 1

EXECUTIVE SUMMARY



Over 80% of content moderators said their employer needs to do more to support their mental health (80.7%).

UNI Global Union survey

Every second, vast amounts of user-generated content floods digital platforms worldwide. Among this content is material that is violent, disturbing or otherwise harmful, and it must be identified and removed to maintain safe, accessible online spaces. We take for granted that we can browse the internet freely, finding what we need without being bombarded by graphic violence, hate speech or explicit content. But without content moderators, the internet would be like a city without law and order – where every street is filled with chaos, danger, scams and disturbing imagery, making it impossible to navigate safely or find what you’re looking for. Content moderators are the invisible frontline workers who shield us from this digital chaos, yet they do so under conditions that expose them to severe psychological harm. Unlike emergency responders and other trauma-exposed professionals, they frequently lack even the most basic workplace protections necessary to safeguard their health and wellbeing.

UNI Global Union and our affiliates have long engaged with content moderators over their working conditions, but for the first time, we have conducted a systematic, global effort to document their experiences. We surveyed 206 moderators worldwide and conducted in-depth interviews with

over forty workers across six countries to develop comprehensive guidelines for making this critical work more sustainable. This is the first study of its kind – and the findings are alarming. Our findings expose a hidden workforce enduring precarious contracts, low wages and psychological distress. We also draw from the research of Equidem to highlight the systemic failures of platform companies in protecting the mental health and labour rights of content moderators.¹

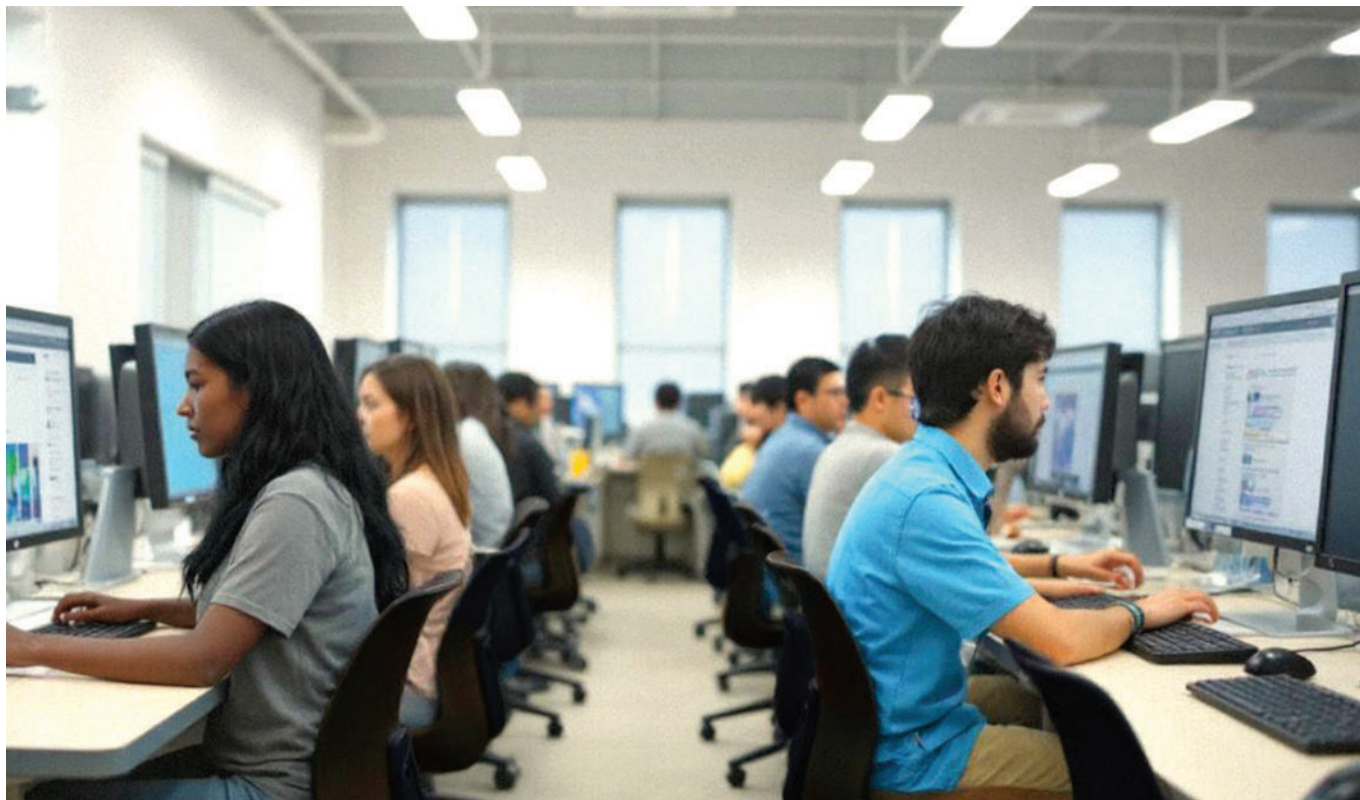
The content moderation industry is structured to distance major tech firms – such as Alphabet, Amazon, ByteDance, Meta and Microsoft – from direct responsibility. These companies outsource moderation to Business Process Outsourcing (BPO) firms like Accenture, Concentrix and Telus Digital and Teleperformance which concentrate jobs in lower-wage countries and regions around the world.

“They tell us to increase speed and they punish us for not meeting the quality standard, which changes every month - and they always take away our bonuses.”

Content moderator, Colombia

Automation is frequently cited as the future of content moderation, yet AI remains deeply reliant on human labour. Moderators train AI systems, review their errors and must identify the most disturbing content – often under increased strain from algorithmic management systems. Rather than replacing human oversight, AI exacerbates the burden on workers, making robust labour protections more critical than ever.

¹ *Scroll. Click. Suffer: The Hidden Human Cost of Content Moderation and Data Labeling, May 2025. Equidem*



To address these urgent issues, UNI Global Union, our affiliates and workers through the Global Trade Union Alliance for Content Moderators have developed eight protocols for safe content moderation that establish essential safeguards for moderators. We are calling on tech companies to implement these practices across their supply chains. These include:

1. **Exposure limits:** Strict time limits on moderating harmful content; use of blurring, grayscale and excerpts to minimize exposure. Mirror best practices in emergency response professions.
2. **Revised productivity metrics:** Eliminate unrealistic quotas and ensure human oversight in performance evaluations. Remove productivity targets for moderating egregious content.
3. **Comprehensive and transparent health and safety training:** Mandatory and re-occurring, trauma-informed training for moderators and their supervisors.
4. **Employee assistance and long-term support:** 24/7 employee assistance access; counselling; critical incident reporting; and trauma-informed resources. Post-contract support for at least 24 months to address delayed-onset trauma.
5. **Stable employment and living wages:** Formalized employment with permanent contracts, living wages and full employment benefits.
6. **Joint occupational health and safety committees:** Democratically elected committees to enforce workplace protections and perform regular audits with transparent health and safety assessments.
7. **Migrant worker protections:** Policies that prevent exploitation, provide job security, protect against labour abuses and support vulnerable workers.
8. **Right to organize a union and collective bargaining:** Respect for the right to organize without interference or opposition from the employer, alongside a policy to allow access to the union. Collective bargaining with recognized unions.

SECTION 2

CONTENT MODERATION: THE HIDDEN WORKFORCE KEEPING THE INTERNET SAFE

Content moderators play a vital role in ensuring safe and accessible digital spaces for billions of users on platforms like Facebook, TikTok and YouTube. Every second, internet platforms generate massive amounts of user content – some of it harmful, illegal or in violation of platform policies. Content moderators are the invisible frontline, sifting through this material to keep online spaces safe. They remove hate speech, misinformation and graphic content, ensuring compliance with both legal and community standards. This work demands constant judgement, emotional resilience and cultural sensitivity, all performed under intense time pressure and strict performance targets.

like debriefing sessions, stress management training and access to mental health care. Content moderators, however, are left to cope alone, with insufficient occupational safety and health (OSH) training or psychological support, despite performing emotionally intensive labour under constant surveillance and time pressure.

In speaking to moderators from across the world, we found that most tend to be hired with low pay and minimal rights. Where workers move countries, employers sometimes house employees in company-chosen accommodations that workers pay for. Many depend on their employer for work

“There is an exclusive Facebook messaging line that handles everything related to inappropriate interactions with children, and it is extremely intense.”

Content moderator, Colombia

To meet this global demand for content moderation, companies rely on a hidden workforce, including many migrant workers who bring essential language skills, cultural knowledge and digital expertise. Digital platforms require culturally specific moderation services for every corner of the world, and these workers are often recruited from neighbouring countries to fill the need.

Despite being exposed to egregious content daily, including violence, exploitation and abuse, many content moderators work without the basic protections given to other trauma-exposed professionals. First responders and emergency dispatchers, for example, benefit from laws that address psychosocial hazards, as well as structured support

permits, that in some cases leave them unable to access public services and protections. These conditions, combined with legal and structural barriers, make it extremely difficult for content moderators to organize collectively or join unions, especially in jurisdictions where migrant worker protections are weak.

A more responsible approach is needed, one that values the expertise of content moderators, provides fair wages, ensures job security and upholds fundamental rights. By recognizing the contributions of this workforce and implementing strong labour protections, the industry can create a safer working environment for those who protect our digital space.

THE BUSINESS OF CONTENT MODERATION: OUTSOURCING ACCOUNTABILITY

Content moderation is a vital operation that keeps digital platforms usable, but instead of handling it in-house, major tech companies typically offload the work to third-party firms – Business Process Outsourcing companies (BPOs). This creates a complex, layered system where global platforms delegate content moderation to outsourcing firms, distancing themselves from direct accountability.

The biggest players requiring content moderation include:

- Alphabet (*YouTube, Google Search Engine, Maps*)
- Amazon (*Twitch, Amazon.com*)
- ByteDance (*TikTok*)
- Meta (*Facebook, Instagram, WhatsApp*)
- Microsoft (*LinkedIn, Xbox Live, other platforms*)
- X (*formerly Twitter*)

While BPOs may hold the contracts, the core aspects of content moderation – such as training protocols, review policies, moderation tools and performance metrics – are largely determined by the tech companies themselves. This arrangement allows platforms to maintain operational control while distancing themselves from direct responsibility for working conditions. In several cases, when workers have raised concerns or attempted to organize, the response has included threats to shift contracts to other jurisdictions, limiting their ability to advocate for improved standards



SECTION 3

FOUR KEY CHALLENGES FACED BY CONTENT MODERATORS

1. EGREGIOUS CONTENT AND VICARIOUS TRAUMA

The prolonged exposure to graphic, violent and hateful material has been associated with significant mental health impacts for many content moderators, including symptoms of post-traumatic stress disorder (PTSD), anxiety, depression and, in extreme cases, suicidality. Despite the well-documented psychological risks inherent in this work,² access to employer-provided mental health support is often limited or inconsistent, leaving many workers without adequate resources to manage the toll of their responsibilities.

“I’ve watched thousands of videos and many types of content. Because of some of them, I have become numb, because of some of them, I have become more profane, and because of some of them, I have nightmares. This job left permanent trauma in my brain.”

Content moderator, Türkiye

Content moderation involves more than reviewing individual posts – it requires confronting some of the most disturbing aspects of human behavior on a daily basis. Moderators interviewed by UNI Global Union described training practices that, rather than mitigating harm, sometimes exacerbated psychological distress. In certain companies,

Will AI replace content moderation?

Automation in content moderation has significant limitations. Despite advances in AI, human content moderators remain essential – and will continue to be – for several key reasons. AI is not autonomous. It depends on human labour at every stage – workers train the models, refine outputs and intervene when automation fails. Human-labelled data is the foundation of every algorithm, and behind the scenes, moderators still make critical decisions about harmful or borderline content.

AI struggles with context, sarcasm, cultural nuances and emerging trends in harmful content. Even the most sophisticated models make mistakes, requiring human review to prevent the spread of egregious material. Tech companies recognize this fact and so they continue to rely on thousands of human moderators to ensure platform safety.

*Moreover, the integration of AI does not eliminate moderation jobs – it transforms them. Instead of removing human oversight, AI often increases the cognitive burden on workers, who must specifically label content and deal with more complex cases AI cannot process. **Far from replacing moderators, as discussed in this report, AI risks intensifying their workload, making workplace protections even more critical.***

The hidden labour behind AI remains largely unrecognized in mainstream discussions, which tend to focus on technological breakthroughs rather than the human workers making them happen. Without content moderators, AI systems would not function effectively. As long as digital platforms exist, there will be a need for human judgement, ethical oversight and safeguards that only workers - not algorithms - can provide.

² Rani, Uma, Morgan, Williams and Nora Gobel. Forthcoming. *The Human Cogs in the AI Machine: Experiences of Data Annotation and Content Moderation Workers in the BPO Sector in India and Kenya*. ILO Working Paper.

“The workload is not the same between all the workers, and some of us get to do three or four times more content than others. For example, I usually work the evening shift and there is a lot more content to moderate but less moderators working, while on the morning shift there are a lot of moderators but less content. When I try to call attention to this problem, I’m being told that this is why they have the good workers in the evening and that we can handle it.”

Content moderator, Portugal

errors in content decisions triggered a requirement to rewatch the same disturbing material multiple times until the correct judgement was made. This approach, intended as a learning tool, instead risked compounding trauma and creating harmful patterns of repeated exposure.

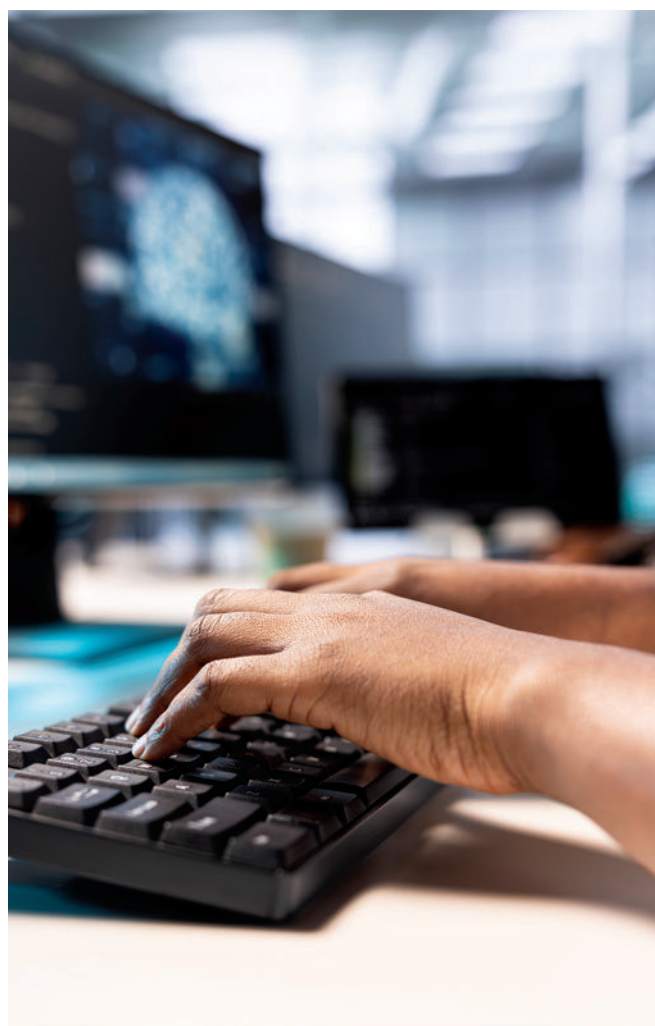
Another moderator reported that one team leader assigned all sexually explicit content to a single worker, believing they were the most skilled at this type of moderation. As a result, the worker was subjected to nonstop exposure to graphic material, with no rotation or reprieve – intensifying the psychological burden and removing any opportunity for relief or recovery.

One moderator described the slow erosion of their humanity:

“We become numb. There is no difference between watching someone eat and watching someone slice open their own arms – because after a while, neither triggers any reaction in us... Beheadings, child abuse, people making love – how many times can you watch these things before you stop feeling anything at all? How many masturbation videos can you sit through before you’re sick to your stomach? How many slurs against women and children can you hear before they stop registering? We watch. We listen. We see. It’s our job. But we lose our ability to be human. We lose our emotions. We can’t be happy. We can’t be sad. We become robots.”
Content moderator, Türkiye.

Some platforms have introduced the blurring of egregious content to protect workers from the psychological effects of harmful material. However, workers said they had to remove the blurring tool when labelling images as they are unable to specifically see the material in the necessary

detail, meaning their performance measures dropped. The pressure to meet the performance targets meant that the blurring tools are frequently not used by the workers we spoke to. Workers believe they are training AI algorithms to moderate content, meaning that the labelling of images must be extremely precise and comply with extensive client policies. We did not speak to workers who had been told by their employers the purpose of the data labelling.





BPO workers in the Philippines demand the right to organize.

Two content moderators we spoke to said the moderation platform they used had an option to switch off egregious content, allowing them to avoid harmful content from time to time. However, this often meant that they received little or no tasks. As a result, they were considered inactive and received no pay for that period.

2. MANAGED BY MACHINES: THE RELENTLESS PACE OF CONTENT MODERATION

Content moderators don't just work – they race against the clock. Their performance is measured by rigid key performance indicators (KPIs) and often managed by algorithms that demand nearly impossible levels of speed and accuracy. One moderator described the relentless pace:

"In just one year, our daily video targets more than doubled. We have to watch videos running at double or triple speed, just to keep up. There's no time to think. No time to process. The only way to hit the numbers is to skip toilet breaks, meals and rest."

Content Moderator, Tunisia

In one site, we heard from moderators who are required to monitor two content streams simultaneously, processing multiple videos at once. Many reported watching videos at double or triple speed, with no breaks between them for hours on end. The pressure to meet ever-increasing targets forces some to work beyond their scheduled hours, sacrificing rest and well-being just to keep up.

One concerning trend that emerged in our interviews is the heavy reliance on performance-based pay. Many moderators reported that up to 50 per cent of their total compensation came from productivity bonuses. These bonuses are not optional perks, they are essential to earning a livable income. This model creates a powerful incentive to prioritize output over personal health and well-being. In practice, it pushes workers to process large volumes of egregious content rapidly, heightening both the risk of burnout and exposure to trauma.

3. LACK OF TRAINING AND SUPPORT

Content moderators are thrown into highly traumatic work often with insufficient health and safety

training,³ leaving them unprepared for the emotional toll of their jobs. Many describe being given only a basic orientation before being expected to handle distressing content – such as violent acts, exploitation and hate speech – on a daily basis.

While some companies provide access to psychotherapists, these professionals are often limited to crisis intervention and triage rather than comprehensive care.⁴ One moderator shared,

“We can book a session, but the therapist is only there to assess if we’re still fit to work. It’s not real support.”

To address this, companies must ensure access to licensed, independent mental health professionals who are specifically trained in trauma and crisis response. These therapists must operate outside the management structure to build trust and ensure confidentiality, so that moderators can seek help without risking their jobs or careers.

4. BARRIERS TO UNIONIZATION

The right to organize in a trade union is recognized as a human right, and the freedom to join a union and collectively bargain is one of the five fundamental principles and rights upheld by the International Labour Organization (ILO). The ability to organize is considered an “enabling right” because it allows workers to collectively defend and advance their other rights through being in a union. If content moderators were able to organize, they could ensure their work is done safely and improve their working conditions.

Canadian company Telus Digital provides services for Google, Meta and TikTok, and identifies employees unionizing as a risk to the business in its annual report.⁵ In Türkiye, when Telus Digital content moderators attempted to organize they were dismissed according to media reports,⁶ while in Germany a union member who spoke to

the Bundestag was suspended by the company only to be reinstated following legal action by his union.⁷

Employing a total of 450,000 workers around the world, California-based Concentrix is one of the largest business service companies in the world, which also provides content moderation services for tech clients. While Concentrix does not explicitly list unionization as a risk to its business like Telus Digital, it does admit that “except for a small number of our employees in certain countries, generally required by local regulations or brought in through acquisitions, our employees are not represented by a labour union, nor are they covered by a collective bargaining agreement”.⁸

“Wages are extremely low and the company fires content moderators if they want to unionize,”

Content moderator, Türkiye

Teleperformance, which employs 490,000 workers around the world, is a major provider of content moderation services and has taken a different approach to Telus Digital and Concentrix. Teleperformance signed a global agreement with UNI Global Union at the end of 2022.⁹ In the agreement, the parties commit to working together to uphold labour rights, including freedom of association and collective bargaining, and ensure a non-oppositional environment for workers to exercise their right to organize. As a result, Teleperformance content moderators in Colombia have organized and secured a collective agreement, while moderators in Portugal and Kenya are in the process of implementing the agreement and organizing with national unions in their countries.

3 ‘Appropriate training’ is a fundamental right at work. ILO C155, Article 19(d)...

4 Comprehensive occupational health and safety services with a preventive focus (see ILO Convention 161 which highlights specifically ‘preventive’ functions. https://normlex.ilo.org/dyn/nrmlx_en/f?p=normlex-pub:12100:0::no:12100:p12100_instrument_id:312306:no.

5 Telus Digital Annual Report 2024, p.16

6 “TikTok workers who watched abusive and traumatic videos say they were fired after trying to unionise”, *The Independent*, 15 March 2025.

7 “Bedeutender Erfolg für Content Moderator”, *ver.di union*, 26 October 2023. “Content-Moderator bekommt Recht”, *ver.di union*, 23 October 2023. “Content moderator in Germany put on leave after testifying over work conditions”, *Reuters*, 22 June 2023.

8 Concentrix Form 10K 2024, p.10.

9 <https://uniglobalunion.org/workers-rights/global-agreements/>

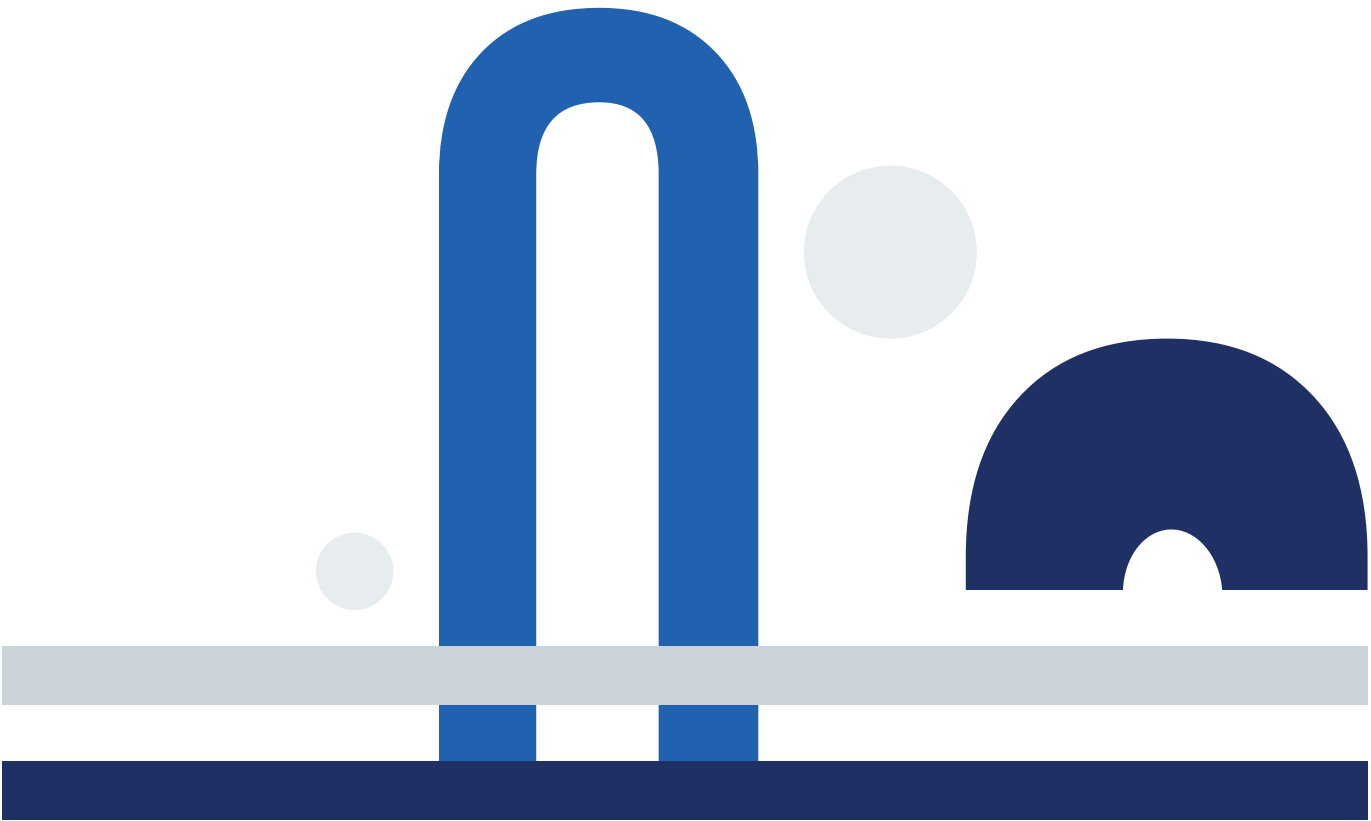
CASE STUDY: HOW OTHER TRAUMA-EXPOSED PROFESSIONS PROTECT WORKERS

While content moderators endure graphic and distressing material daily, they lack the workplace protections afforded to other trauma-exposed professions. Here's how different sectors address psychological hazards:

Profession & location	Protections
First responders (paramedics and police services) Canada	<p>Mandatory psychological debriefing after critical incidents, peer support programmes and access to long-term mental health care. There are many mental training programmes available to personnel, including critical incident stress management (CISM), critical incident stress debriefing (CISD), peer support, mental health first aid and Road to Mental Readiness (R2MR). These exist to minimize the impact of stressors on personnel.</p> <p>Critical Incident Stress Debriefing (CISD): After exposure to a traumatic incident, workers are debriefed within 24-72 hours with trained counsellors and/or peer support groups.</p> <p>Regular mental health screenings: Many departments offer regular psychological evaluations, covered by employer insurance.</p> <p>Workload caps and rotations: To prevent chronic exposure, some departments rotate staff out of high-trauma assignments (e.g., homicide or child abuse units) after a set time.</p> <p><i>Source: Canadian Institute for Public Safety Research and Treatment (CIPSRT), 2022</i> https://www.cipsrt-icrtsp.ca</p>
Emergency dispatchers USA, Germany	<p>Stress management training, rotational shifts to prevent burnout and counselling services covered by employer benefits.</p> <p>Post-traumatic call procedures: After handling traumatic calls, dispatchers may participate in critical incident stress debriefing (CISD) sessions. CISD is a structured process aimed at mitigating stress and promoting recovery by allowing emergency personnel to discuss their experiences and emotions in a supportive environment. There are no productivity targets for dispatchers who have returned to work after a critical incident.</p>
Journalists covering conflict & crisis (Reuters, BBC, AFP) Global	<p>Trauma training, mental health check-ins and access to specialized therapists. Some organizations offer time off after distressing assignments.</p> <p>Hostile Environment & First Aid Training (HEFAT): Training includes trauma awareness, stress responses and psychological self-care before field assignments.</p> <p>Manager support: Editors are trained to monitor for signs of trauma and can reassign journalists covering distressing topics (e.g., child abuse, war crimes).</p> <p><i>Source: Dart Center for Journalism & Trauma, Columbia Journalism School</i> https://dartcenter.org</p>
Child protection workers Sweden, Netherlands	<p>Regular mental health assessments, case reviews to reduce exposure to high-risk content and structured support networks.</p>

KEY LESSONS FOR THE CONTENT MODERATION INDUSTRY:

Best practice	Relevance to content moderation
Pre-task psychological training	Equip moderators with resilience tools and normalize mental health conversation
Mandatory mental health debriefs	Compulsory post-exposure support, especially after viewing severe content
Workload caps / trauma rotations	Implement structured limits on exposure to egregious material
Access to professional counselling	Offer free, confidential and culturally competent mental health services
Institutionalized duty of care	Embed worker mental health into operational models – not as an optional add-on



SECTION 4

EIGHT PROTOCOLS FOR SUSTAINABLE CONTENT MODERATION

Using our survey data and interviews capturing input from hundreds of content moderators along with our research into protections of other trauma-exposed workforces, we have developed eight concrete protocols for sustainable content moderation.

1. Limit exposure to egregious content

Workers' exposure to egregious content should be limited to one hour per day, five hours maximum per week – pairing this cap with regular task rotation to less graphic or demanding roles. To minimize harm, content should be blurred, shown in grayscale, or presented as excerpts rather than full clips. This mirrors best practices in emergency response professions, which limit exposure to traumatic incidents to prevent burnout, PTSD, suicidality and suicide. If a moderator finds a specific queue traumatic or triggering, they should be able to request reassignment to a different queue without facing any penalties or repercussions.¹⁰ Long-term exposure to less extreme, but still disturbing content, can have a cumulative impact, especially if no limits are placed on working hours.

Over 70% of respondents moderate “sexually explicit or violent content”

UNI Global Union survey

2. Revise productivity metrics with human oversight

Tech companies must replace rigid, automated KPI systems with human-led performance evaluations that account for the mental health and well-being of content moderators. Productivity expectations should be realistic and adjusted based on the

complexity of the content being reviewed – with no quotas for moderating egregious content, eliminating KPIs that encourage workers not to use blurring tools. Workers must have the right to challenge algorithmic decisions about promotions, pay and termination, with a fair appeals process and representation. Productivity targets must be transparent and communicated to workers during hiring so that they are informed of the service levels and can make informed decisions. No worker should have to skip breaks to meet quotas – safe staffing levels must be enforced. Companies must also redefine “working time” to include basic human needs, so moderators are not penalized for reasons such as bathroom breaks or rest periods.

3. Comprehensive health and safety training for moderators and managers

Research demonstrates a clear connection between health and safety training and positive outcomes for workers. Employers must ensure full transparency during recruitment, providing detailed information about the role to enable moderators to make informed decisions and implement harm-reduction practices from the start. Job descriptions should explicitly state that the role involves moderating sensitive, explicit and potentially harmful content, with generic examples provided during interviews.

Before engaging with sensitive content, moderators must complete mandatory training during work time, including culturally sensitive, trauma-informed, psychological health and safety training, repeated annually and incorporated into paid work hours. Training should cover wellness programmes, psychological coping strategies and procedures for requesting alternative work placements. Employers must collect a signed acknowledgment confirming that moderators understand available resources and policies.

Managers and supervisors must complete mandatory trauma-informed training within two weeks of

¹⁰ Note: workers have a right to refuse without fear of ‘undue consequences’ (ILO C155, Article 13) if they believe their work presents a ‘serious and imminent’ risk to their ‘life or health’ (ILO C155, Article 19(f)).

onboarding, focusing on supporting moderators' well-being and managing employees in distress, with paid annual refresher courses.

Non-disclosure agreements (NDAs) are frequently introduced during initial training sessions and often function to silence workers and discourage them from raising concerns about their working conditions. This practice creates a chilling effect that undermines transparency and accountability. While NDAs may be used to protect proprietary information or client content, they must not extend to issues related to occupational safety, mental health or general working conditions. Companies must ensure that NDAs do not interfere with a moderator's ability to report harmful practices, organize for safer conditions, or seek support.

When respondents reported having been adequately trained, complaints about mental health issues were cut in half.

UNI Global Union survey

Workers should be clearly informed during their training sessions that NDAs do not apply to discussions about their well-being, workplace treatment or union rights.

4. Employee Assistance Programmes and long-term support

Companies must provide comprehensive Employee Assistance Programmes (EAPs) with 24/7 support, counselling, family assistance and critical incident response. Licensed, independent therapists with expertise in vicarious trauma must be available during paid work hours. Workers should have a clear, accessible pathway to report mental health concerns without fear of job loss or discrimination, while protecting worker confidentiality. Peer support groups should meet at least monthly with employer-provided resources. Individual and group counselling must be available, providing moderators with at least 45-minute individual sessions available per month. A crisis-trained licensed therapist must be available whenever moderators work.

To build trust, therapists must not only be professionally qualified but also demonstrably independent of the employer. Many workers fear disclosing mental health struggles due to concerns about

job security or retaliation. Companies should fund independent therapists who operate through or are co-managed by the union, ensuring confidentiality and psychological safety.

Wellness resources, including EAP access, must continue for at least 24 months post-contract, ensuring long-term psychological support for delayed-onset trauma.

5. Guarantee long-term, stable employment contracts and living wages

Content moderators' employment contracts should have full-time, permanent positions with benefits such as health insurance, paid time off, paid sick leave and career development opportunities. Providing long-term, stable employment is essential to valuing both the job and the people who do it. Like other professions where trauma is inherent, content moderation is a profession and must be recognized as such. Furthermore, implementing living wages is crucial to ensuring that content moderators are not only fairly compensated for the emotional and psychological toll of their work but also able to support themselves and their families with dignity. Exactly how much performance bonuses represent relative to base pay depends on the country, but our study found examples where half of the moderator pay is linked to performance-based bonuses tied to speed and output. Through making these bonuses essential, workers are pressured to push through traumatic content to earn a dignified wage.



6. Democratically elected and trained occupational health and safety committees

Joint occupational health and safety (OSH) committees, where workers elect their own representatives, are essential to identify risk, prevent harm and effectively implement health and safety programmes. Their importance is reinforced by international labour standards: ILO Recommendation 164 affirms workers' rights to safety representation and equal participation, while ILO Convention 155 highlights the need for robust occupational health and safety frameworks for all workers.

Protecting content moderators requires more than one-time interventions – it demands sustained efforts such as periodic audits, collecting worker feedback and tracking health outcomes. Trade unions and OSH committees are central to these processes and ensure that worker perspectives are reflected in both the design, oversight and continuous improvement of safety measures.

7. Protect migrant workers

Migrant workers are a crucial part of the content moderation workforce, yet they face unique vulnerabilities.¹¹ Tech companies must ensure that BPOs provide migrant moderators with fair and secure working conditions, including offering employment contracts in the worker's native language so that workers fully understand their rights, responsibilities and terms of employment. Additionally, employer-funded legal assistance for visa and work permit applications is essential for compliance with local labour laws. Where relevant, work contracts should not tie legal residency or work permits exclusively to the employer, preventing situations of forced dependency.

Employers should provide safe, dignified and independent housing options, rather than company-controlled accommodations that restrict movement and autonomy. Companies should support

migrant workers to choose their own accommodation, where the rental agreement is independent of their employer.

Migrant workers must have the freedom to join unions without fear of retaliation linked to their visa status. Employers must inform workers in writing of their right to join a union, and this step is even more important for migrant employees who have greater vulnerability than those who are contracted in their home country.

Access to professional mental health and community support is also crucial, with regular individual and group sessions facilitated by trained and certified mental health professionals to address the unique stressors facing migrant workers. Employers should conduct wellness checks if a migrant worker is absent from work and notify the relevant authorities if they do not respond. Furthermore, BPOs should support new workers in building social and professional networks, helping them integrate into their workplaces and local communities.

8. Respect the right to organize a union and collective bargaining

Freedom of association is crucial for the protection and empowerment of content moderators and other tech workers worldwide. By joining a union, workers collectively empower themselves to ensure their rights are upheld, their voices are heard and their well-being is prioritized now and into the future. BPOs should recognize unions without interference or opposition, participate in collective bargaining, and engage in transparent and constructive dialogue to address workers' concerns. This engagement is foundational to create safer work and it contributes to the long-term success of the industry by promoting ethical practices and improving employee well-being and retention.

¹¹ *International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families* – <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-convention-protection-rights-all-migrant-workers>

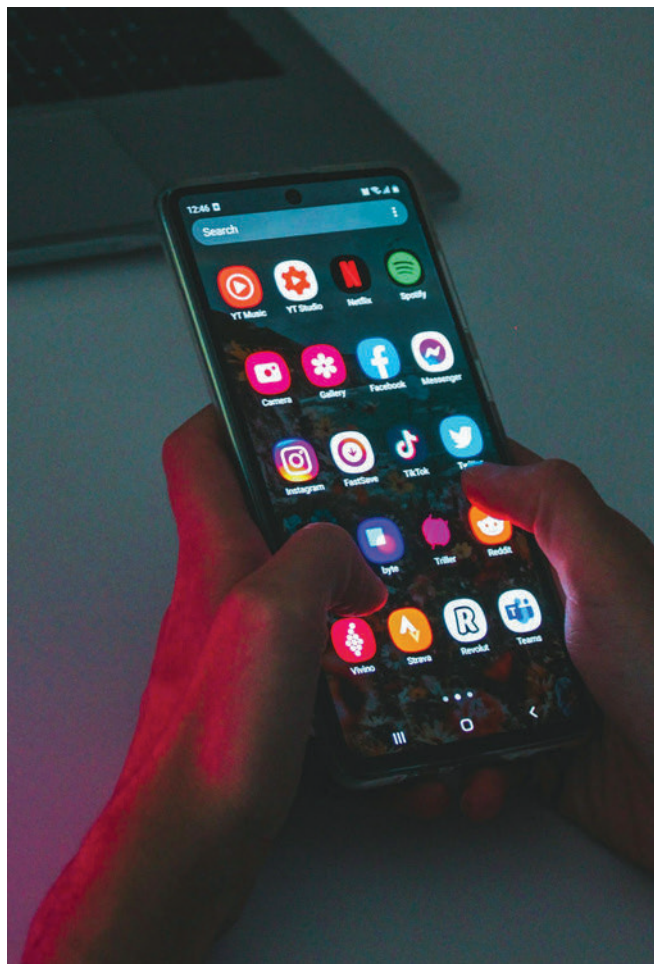
CONCLUSION: A CALL TO ACTION

Many content moderators take deep pride in their work – protecting children, shielding fellow citizens from harm, and upholding the integrity of digital public spaces. Like first responders, they perform difficult and often invisible work for the public good. They deserve the same dignity, recognition and protection we afford to others doing society’s most vital jobs.

AI can support moderation, but it cannot replace human judgement. Only people can navigate the cultural, linguistic, and ethical complexities needed to assess harm and context.

For the first time, tech firms and the BPOs in their supply chains have clear, evidence-based guidance on how to achieve this: by adopting the Eight Protocols for Sustainable Content Moderation. At the center of these protocols are freedom of association and collective bargaining – tools that empower workers to shape the conditions under which they labour.

If there is no action, then we – as users of these platforms – must raise our collective voices to stand with the people who shield us from the unimaginable. It’s time to stand with the people behind the screens.





8-10 Avenue Reverdil
1260 Nyon, Switzerland
www.uniglobalunion.org