

The People Behind Safer Screens

Public briefing on the content moderators (CMs) who help ensure online safety

HOW MANY CONTENT MODERATORS ARE NEEDED?

There is no single transparent global register of CMs. Much of the workforce remains hidden behind outsourcing chains and limited public reporting but it is a large and internationally dispersed workforce.

~15K

Meta CMs

~10K

Google/YouTube CMs

Global Workforce

U.S. • Philippines • India • Ireland • Kenya • more

A complex workforce

CMs vary by geography, language, employment status and life circumstances. These conditions can shape how risk, support and safety are experienced.



HUMAN INVOLVEMENT IS CRUCIAL TO ONLINE SAFETY

Human judgment remains essential when content or context is complex. A range of categories are reviewed:



Illegal & highly harmful

Terrorism, child sexual abuse material (CSAM) fraud, violent crime, illegal drugs/weapons, self-harm content



Community standards

Hate speech, bullying, harassment, graphic content, impersonation, spam, misleading content



Context-dependent

Intent, repetition, threats, exploitation, target identity and wider context



Legal but harmful

Borderline material that may create risk, especially for children



Child-safety cases

Grooming indicators or cases requiring specialist review/escalation

WHY PUBLIC AWARENESS MATTERS



What Public Awareness Promotes

- Understanding of the need for safety measures**
Most people do not realize the amount of harmful content that is posted online.
- Appreciation of the role of content moderators**
Most of us do not see the worst material because CMs reviewed and removed it.
- Understanding the psychological cost of the work**
Empathy for the toll this takes on CMs is needed for changes to be made.
- Shared responsibilities**
Awareness and empathy can lead to advocacy for prevention of harm to CMs by employers and policymakers.



The Human Cost

Because this work is often repetitive, high volume and fast paced, the cumulative impact of reviewing harmful material over time may be as important as any single piece of content. Repeated exposure to disturbing content is associated trauma-related reactions such as:

Nightmares

Depression

Anxiety

Anger

Physical symptoms

Intrusive thoughts

Isolation

Insomnia

Support helps, but wellbeing services are not a substitute for reducing preventable exposure and harmful job demands.

A SAFER INTERNET REQUIRES SAFER WORK

Greater awareness should lead to better questions about how platforms and vendors design and support moderation work.

Realistic expectations

Manageable workloads

Exposure limits

Adequate breaks

Peer support

Confidential care

Safe conditions

TRANSPARENCY MATTERS

Tracking exposure, turnover, sickness absence, psychological incidents and access to support could help show whether protections are improving in practice.

