

Powering digital defense through an AI-plus-human approach



User-generated content (UGC) is prolific

In April 2022:



1.7 million
pieces of content were shared on Facebook



500 hours'
worth of video were uploaded to YouTube



347,000
tweets were posted every minute

Source: MIT Technology Review Insights, *Humans at the Center of Digital Defense*

For usable online experiences, **content moderation** — the monitoring of UGC — is essential.

Content moderation is more than social media monitoring



Any site that allows information to come in that's not internally produced has a need for content moderation.

— **Mary L. Gray**, senior principal researcher, Microsoft Research



30% of large companies will consider content moderation a top priority by 2024
(Gartner)

What do content moderators do?



Review content where context is critical



Monitor for culturally sensitive content



Address phony reviews and ratings



Police deceptive advertising



Ensure safe online interactions among users



Validate content flagged by AI



Human guidance is key

Refine training data for ML solutions by flagging content that escapes the algorithm

Identify context or navigate cultural, regional, and social intricacies

Provide accuracy and context for emerging types of digital interactions

Technology enables content moderation at scale

Machines can dramatically reduce the massive volume of content that needs human oversight.

Automated, API-enabled natural language processing (NLP) algorithms detect harmful written content.

99.1%

automated proactive detection and removal of both terrorist content and violent and graphic content
(Facebook)

Computer vision technology monitors images and visual content.

Up to **99.7%**

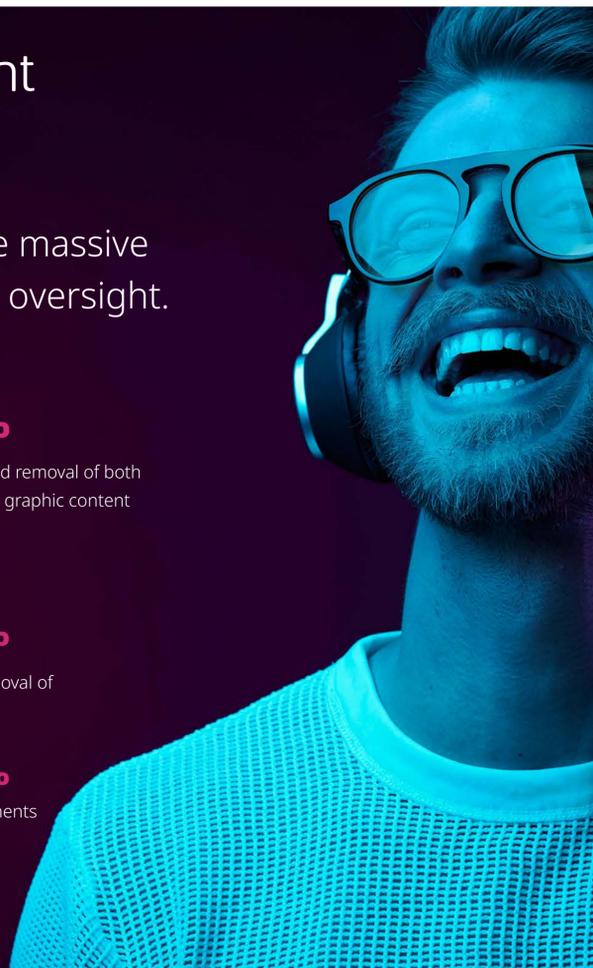
automated proactive removal of fake accounts and

Other content moderation AIs flag anomalies in user identity and account verification scenarios.

98.9%

of inappropriate comments
(Teleperformance)

Speech-to-text models handle video and spoken content.



Learn more about the **synergy between humans and technology** in today's digital defense landscape.

Download the latest white paper from MIT Technology Review Insights, in partnership with Teleperformance.