



Take a holistic, humane approach toward content moderation for brand safety

JANUARY 2023

Ashish Chaturvedi, Practice Leader

A real-time content moderation strategy has become a prerequisite for every online platform. Approximately 4.6 billion users generate more than multi-quintillion bytes of content daily. Tech Mahindra, via its Trust & Safety Services practice, is working with some of the world’s prominent technology and social media companies to overcome the biggest content moderation challenges. As content continues to expand, with new forums such as the metaverse, adding complexity to the moderation paradigm, HFS caught up with Tech Mahindra leadership to learn more about its approach to this sensitive topic.

The “AI + moderator” hybrid content moderation approach is the holy grail, but technology still plays second fiddle

Tech Mahindra’s Trust & Safety Services leadership focuses on achieving consistent delivery quality, solving content moderation challenges, and going beyond “just” making money. To effectively protect their brand,

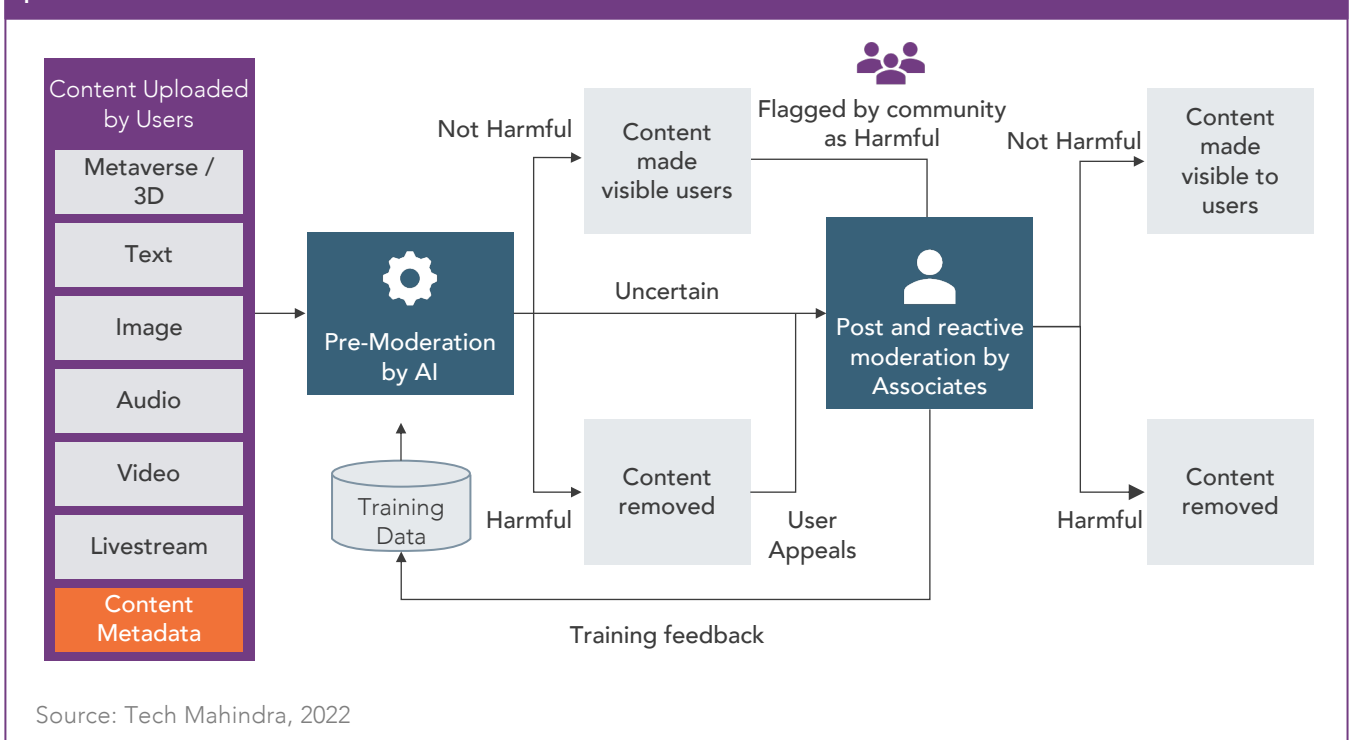
marketing and brand leaders must fully understand the various business aspects—people, technology, delivery ethos, client engagements, and program objectives.

Most of Tech Mahindra’s technology and social media clients have in-house, artificial intelligence-led monitoring platforms as their first line of defense. However, the current technological capabilities are limited. For example, deep learning algorithms cannot wholly avoid false positives, understand the social context, or detect complex emotions like sarcasm. Therefore, humans still manage most processes illustrated in Tech Mahindra’s “AI + moderator” hybrid content moderation methodology in Exhibit 1.

Noteworthy highlights of the “AI + moderator” model include

- **An additional layer of vigilance:** AI makes the jobs of Tech Mahindra’s 60,000 moderators more manageable. Before reaching a moderator, potentially harmful content is reviewed twice, first by AI and then by the community.

Exhibit 1: Crowdsourced vigilance undergirds Tech Mahindra’s content moderation process



Source: Tech Mahindra, 2022

- **Complementary tech support:** Tech Mahindra offers tech add-ons to fully leverage the technological potential of its clients' monitoring systems. For example, it provides audience profiling (user segmentation) models to help the existing monitoring system understand its users, propensities, likes, and dislikes.
- **Vast language coverage:** The team can handle content in over 50 languages, including the elusive ones such as Chinese, Korean, and Japanese. The team has transcribed more than 3,000 hours of audio in multiple languages and accents.
- **The moderators train the AI system:** The moderators are trained to feed inputs into the AI system whenever overturning an original decision, helping the deep learning algorithm become more intelligent.
- **Standardized processes:** There are standard templates for all essential functions, including built-in templates for data collection, annotation, translation, transcription, metadata tagging, and content moderation.
- **A sorted QA framework:** The QA framework has a quick feedback process (typically within 24 hours) and offers outlier management, reporting and analytics, and an automated SME helpdesk.

Keeping the bar high through quality recruitment and moderator wellness program

Tech Mahindra's content moderator candidate screening process, developed in partnership with in-house behavioral psychologists, first evaluates each candidate on multiple dimensions: content moderation and language knowledge, behavioral patterns (via psychometric testing and Patient Health Questionnaire-9 to assess depression severity), scenario-based actions, and prior content moderation experience. The next leg of the screening process takes it a notch

higher, evaluating candidates based on their political views, mental resilience, and cultural knowledge. For example, a candidate with a strong political leaning might have an unconscious bias toward a political group, proving to be a poor fit for political comments moderation work. Such candidates are assigned to other work queues.

It's more complicated to ensure the moderator can deliver consistent quality. HFS Research's extensive work with the service provider community points toward continuous learning (skill development), financial growth, and employee wellness as significant components in moderator productivity. Tech Mahindra leverages multiple training methods, including e-learning, video tutorials, simulations, and gamification. Tech Mahindra briefly discussed its transparent and rewarding moderator career roadmap to leverage financial growth and motivate moderators. HFS found it on par with the competition but not a differentiator. However, its moderator wellness program sets Tech Mahindra apart from the competition.

The wellness program offers a tailor-made module for each moderator based on their current scope of work. For example, moderators exposed to gore, violence, and hate content get 150 minutes daily from their 9-hour shift for wellness. An effective resilience support structure is in place, including in-house wellness coaches, a single point of contact, and dedicated wellness officers. Employees can also access regular health check-ups, a dedicated wellness app, and counselors 24 hours a day. Conducting regular wellness awareness sessions, such as a session on unconscious bias, training managers to gauge warning signs (performance or behavior deviation), and certification programs for wellness coaches are all part of the drill. HFS believes that the focus on moderator wellness is paramount, given exposure to the extreme, potentially harmful content and the mental fatigue associated with real-time content screening.

Where the rubber meets the road: looking at real-world client success

The Tech Mahindra team discussed more than a half-dozen high-profile enterprise clients ranging from one of the top 10 gaming companies to one of the most prominent technology software vendors. A few things stood out:

- **High levels of accuracy, low levels of false positives:** AI + humans help deliver 97% accuracy with less than 1% of false positives across the board.
- **Managing large volumes of content:** Tech Mahindra reviews 15 million ads, appeals, and tickets annually for a technology major.
- **Enterprises collaborating with Tech Mahindra to devise new policies:** The emergence of new engagement models

via concepts such as metaverse is giving rise to unique content moderation needs. Tech Mahindra is working with several clients to decode the space and define new review rules.

With more than 10 large clients, Tech Mahindra's Trust & Safety Services practice has been witnessing double-digit year-over-year growth, generating more than \$50 million in annual revenue. The team aspires to become a purpose-driven business. One of the initiatives to double down on this statement is prioritizing employee wellness over financial profits. HFS believes this is the right approach. Any enterprise needing content moderation services needs to understand and support the challenges inherent in the market, rewarding those providers who balance quality of service with employee wellness.

The Bottom Line: An AI + moderator model is here and now, but the future is leveraging technology to a greater degree while continuing a laser focus on moderator wellbeing to ensure employees are supported to deliver on what AI cannot accomplish (yet).

The explosion of user-generated content and new engagement channels (such as metaverse) is compelling digital enterprises to partner with experienced providers to manage and run sizeable multi-country content moderation engagements. Enterprises looking to address the challenge of content moderation must evaluate providers' investments in both human and technology capabilities.

HFS Research authors



Ashish Chaturvedi

Practice Leader

Ashish Chaturvedi is a Practice Leader at HFS covering supply chain, retail, and disruptive technologies. Ashish is an accomplished IT industry analyst and a member of the IDG Influencer Network. With over 13 years of technology research experience, Ashish has authored more than 65 research reports spanning retail technologies, enterprise modernization, low-code/no-code development, digital benchmarking, and IT sourcing.

Over the years, Ashish has advised several senior executives on digital strategy, product/service planning, next-gen technologies, and IT procurement. He has delivered several multidisciplinary research engagements including provider and market intelligence reports, go-to-market workshops, white papers, podcasts, and research-based advisory.



About HFS

Insight. Inspiration. Impact.

HFS is a unique analyst organization that combines deep visionary expertise with rapid demand-side analysis of the Global 2000. Its outlook for the future is admired across the global technology and business operations industries. Its analysts are respected for their no-nonsense insights based on demand-side data and engagements with industry practitioners.

HFS Research introduced the world to terms such as “RPA” (Robotic Process Automation) in 2012 and more recently, Digital OneOffice™ and OneEcosystem™. The HFS mission is to provide visionary insight into the major innovations impacting business operations such as Automation and Process Intelligence, Blockchain, the Metaverse and Web3. HFS has deep business practices across all key industries, IT and business services, sustainability and engineering.

Read more about HFS and our initiatives on:
www.hfsresearch.com or follow [@HFSResearch](https://twitter.com/HFSResearch)