The New Era of **Content Moderation**

How artificial intelligence (AI) is the helping hand content moderators need to address today's complex digital landscape.
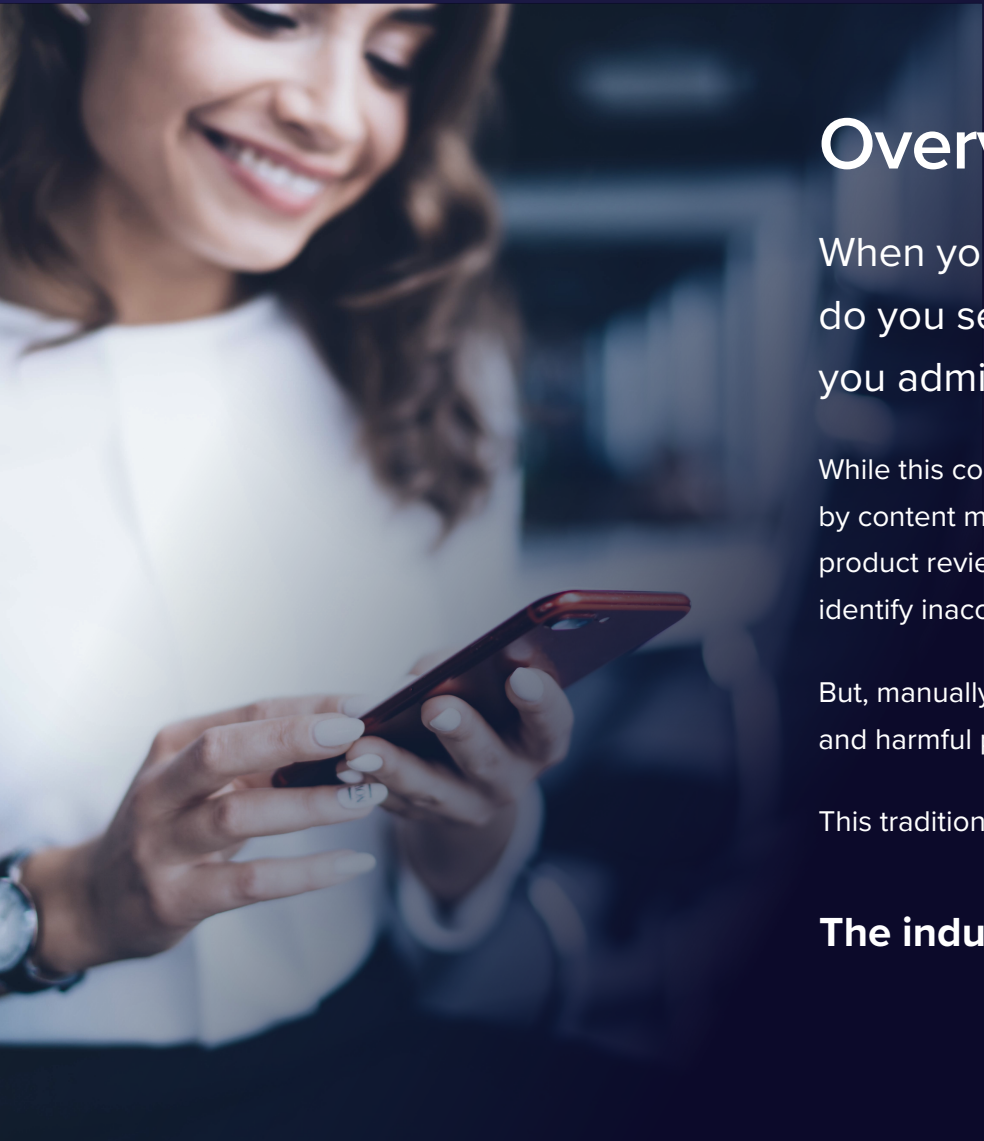
Overview

When you open up your favorite social media platform, what do you see? A feed of content from brands you love, users you admire, and posts you're excited to read.

While this content is carefully curated by brands and users, it's also carefully managed by content moderators — workers tasked with manually reviewing social media activities, product reviews, and other user-generated posts. They siphon through this content to identify inaccuracies and flag malicious or inappropriate posts for removal.

But, manually reviewing each piece of content — as well as reading through malicious and harmful posts — is both time-consuming and psychologically draining.

This traditional approach to content moderation isn't sustainable.

The industry is ripe for innovation.

The Trouble with **Traditional Content Moderation**

As society embraced the information age, digital channels and user-generated content exploded. In response, organizations had to find new ways to cut through the noise and establish highly personalized conversations with online audiences.

Stories surfaced in every shape and size, with the production and management of content growing more complex by the day. Content became 24/7, increasing 10x over the past few years alone.[1]

Naturally, managing and storing all this content became quite the feat. On Twitter alone there are around 500 million tweets sent each day, or 6,000 tweets every second.[2] It's a lot to keep up with and even more to somehow organize, filter, and moderate.

The nature of content itself only compounded this problem. In the digital realm, millions of voices leverage social and digital media to share their two cents. While this leads to unique storytelling, widespread knowledge-sharing, and strengthened connections, it can also lead to the spread of contentious and harmful messages or even misinformation.

10X
Increase in content creation over the past few years alone

500M
Tweets sent each day, or 6,000 tweets every second.

65%
Of harmful or misleading tweets are shared by 10 "fake news" and conspiracy outlets

In fact, George Washington University researchers found that ten predominantly "fake news" and conspiracy outlets were responsible for 65% of tweets linking to such stories.[3]

This is precisely why content moderation is so vital. Content moderation is effectively an act of public good. Moderators are like firemen or first responders, tasked with shielding members of society from harm.

In order to root out this negativity, moderators must look through a diverse spectrum of content. The good, the bad, and the ugly. The type of material – and sheer volume of it – is often overwhelming, which impacts on moderator happiness and well-being.

As moderators increasingly face a mountain of complex content and organizations struggle to protect them or their users, companies must start looking beyond the traditional methods for content moderation.

Social media users have "not too much" or "no confidence" in social media companies to completely remove four different types of content

62%	53%	55%	52%
False information	Hate Speech	Harassment	Offensive content

Deploying the Dream Team:
Human and Machine

Artificial Intelligence (AI) is the helping hand content moderators
need to address today's complex digital landscape.

That's because AI-based platforms continuously analyze and learn to recognize certain patterns
through automated cognitive intelligence, making the moderation process quicker and smarter
each time. However, AI is just that – a helping hand, not a silver bullet. Why is this?

While AI's deep learning capabilities empower it to successfully moderate content such as
images, it can also struggle with text and video, which are highly contextual and nuanced. In fact,
AI only catches 16% of posts involving bullying and harassment on Facebook, and about 80% of
hate speech.[4]

AI is great for proactivity and speed, but the human touch is still required in order to stay ahead of
today's dynamic threats. AI requires the additional context – including societal, cultural, and
political factors – that human moderators provide.

Content moderators bring human empathy and situational thinking to the table, which is why
Sutherland calls them Content Guardians. These Guardians strive to protect public safety and
well-being via content moderation, examining posts through a lens of compassion. And when the
human eye works with AI, the two become a content moderation dream team.

This is the backbone of Sutherland's Content Moderation solution, which blends the strengths of
humans and AI to moderate content at scale to create safe and trustworthy online environments
for organizations and their communities.

SUTHERLAND®

Sutherland's AI-enabled platform automates content screening, while its proprietary Happiness and Social Indexes and dedicated psychologists protect moderators and monitor well-being, empowering them to safely and effectively do their job. This approach shields moderators from danger as they work to do the same for society.

Quality Content, **Better Experiences**

While the digital age continues to evolve, content will remain a constant. Continuous innovation, digital trends, and user behavior have ensured as much.

But, with new challenges comes avenues for opportunity. By embracing proactive, advanced approaches to content moderation, organizations reap tangible benefits, such as time savings, employee retention, and improved brand reputation.

Through effective content moderation, businesses not only get a handle on the influx of digital content, they also future-proof their organization. With proactive moderation approaches and strategic technology in play, companies get to focus on what matters – customer experience.

Outdated processes are a thing of the past. **It's time to pave the way for experiences engineered for the future.**

SUTHERLAND®

Over the past 30 years, we've helped our customers skillfully navigate the hype cycles for every emerging technology while delivering the greatest business outcomes. With Sutherland, you get the best of both worlds: all the technology, tools, and resources of the big players in the digital space delivered in an agile, people-centric company that gets you.

For more information on how we can help you transform your processes, **contact us via our website, email us at sales@sutherlandglobal.com, or call us at 1.585.498.2042.**

→ **Contact Us**

SUTHERLAND®

Resources

1. Steinhour, Jill. *Double Treat - Adobe and Econsultancy: Second Annual Survey of B2B Digital Trends.* Adobe Blog, Adobe, 14 Nov. 2016
theblog.adobe.com/double-treat-adobe-and-econsultancy-second-annual-survey-of-b2b-digital-trends/#:~:text=Jordan%20Kretchmer%2C%20founder%20of%20LiveFyre,the%20right%20person%20at%20the.

2. Twitter Usage Statistics. *Twitter Usage Statistics - Internet Live Stats*
internetlivestats.com/twitter-statistics

3. Zhou, Marrian. *Fake News on Twitter Is Still Reaching Millions, Study Finds.* CNET, CNET, 5 Oct. 2018
www.cnet.com/news/twitters-fake-news-problem-is-still-very-much-active-study-finds/

4. Facebook Transparency Report: Community Standards. *Facebook Transparency Report | Community Standards*
transparency.facebook.com/community-standards-enforcement#bullying-and-harassment.