CONTENT MODERATION:

The Future is Bionic

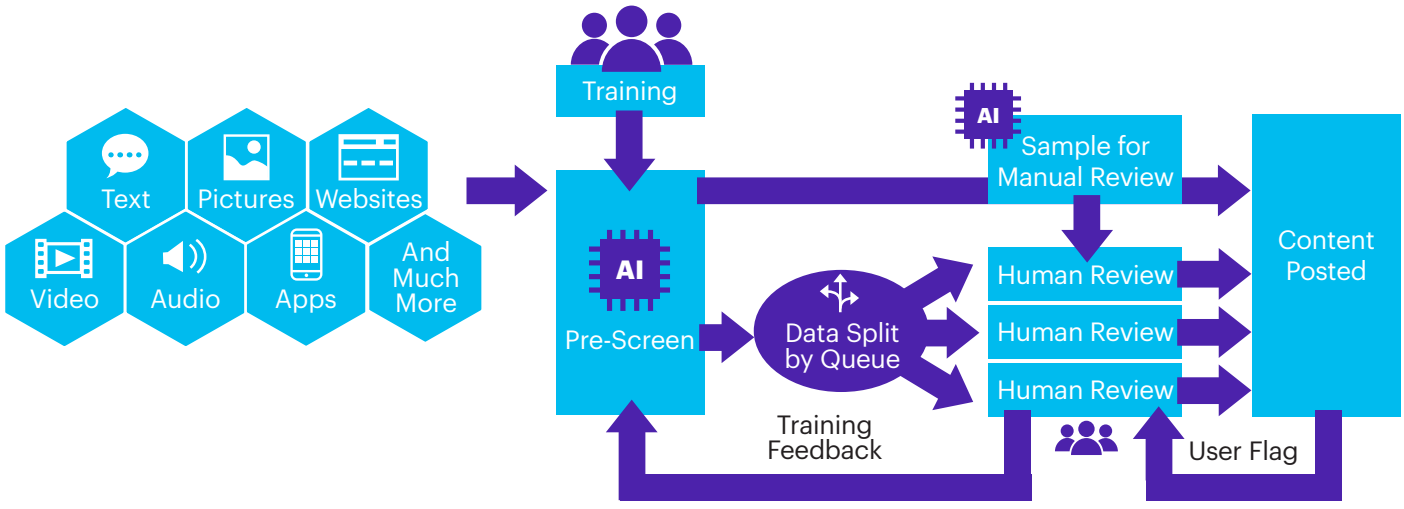
accenture

Is that email offer from a big brand name really a phishing scam? When is a seemingly benign social media post really cyber bullying? In the absence of context, content moderation can be a shot in the dark.

Since the birth of the Web, but especially since the rise of social media, content moderation has become one of the largest and most secretive operational functions in the industry. An army of moderators around the world—in Manila and Bangalore but also in the U.S., Canada and Europe—screen through violence, pornography, hate speech and mountains of other illicit, illegal or inappropriate content. Estimates vary widely, but most indicate that more than 100,000 people are moderating content globally. The work is difficult to stomach. Attrition is high.

Artificial Intelligence offers the promise and threat of being able to take over much of the work. At some social platforms, AI now catches more inapropriate images than people do. Others are working on AI that can understand what's happening in a live streaming video. Others still are working to identify the kind of harassment and bullying that requires contextual understanding rather than just identifying an inappropriate word or image. There's no question that the way companies monitor digital content and remove offensive material is on the brink of change.

Contrary to the headlines predicting that AI will supplant all human moderation, Accenture believes that the future is built on a synergistic relationship between human moderators and AI. This "bionic" model will change both the AI used and the people required. No longer will there be a need for thousands of low-skilled positions to do content review. Instead, a new role will emerge augmented by artificial



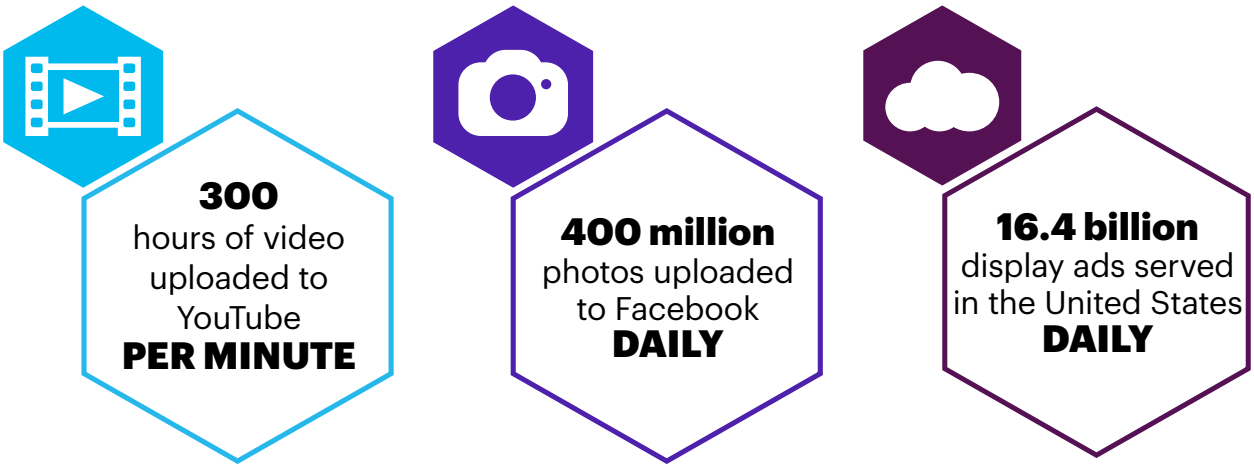
Content is returned securely with approximate action and category identification, and can often be resolved automatically. Manual reviews are also completed for content flagged.

intelligence (AI). Highly skilled content forensic investigators will garner senior positions in Internet-based companies and enjoy lucrative career paths. This transformation will help rapidly growing digital platform providers scale content moderation at an acceptable risk tolerance and at affordable cost.

THE CONTENT MODERATION IMPERATIVE

Every industry is now digital. Retailers are growing their on-line shopping channels, health care companies are managing electronic health records and makers of products from technology to industrial equipment and consumer products are promoting their wares on the web and through digital advertising. Digital platforms are growing exponentially as consumers and companies self-publish content in social media channels at astonishing rates. The increasingly heterogeneous mix of content adds complexity to the review process while the pressure to quickly recognize and respond to unacceptable content intensifies.

These issues are causing stress on the screening protocols and processes of both content producers and platforms, and the associated risks are rising. Internet companies have a responsibility to adhere to diverse global regulations and protect consumers from fraud, spam, phishing, malware, malicious networks and abusive content. A failure by the producer or platform to catch inappropriate content can result in financial risk, brand degradation, loss of consumer trust and personal risk for viewers or recipients. Upcoming regulations, such as the EU's General Data Privacy Regulation, will heighten these challenges, placing even greater responsibility on Internet companies to take actions that protect individual privacies. Current content moderation methods are unsustainable. Something has to change.



Sources: YouTube, Facebook, comScore and Accenture analysis.

A SYNERGISTIC ARTIFICIAL INTELLIGENCE RELATIONSHIP

Many Internet companies are already implementing tools such as image recognition that help with content moderation by identifying objects within images. Weighing factors such as user experience and risk, these tools determine if images should be human-reviewed. This process is eliminating large volumes of content from the investigator's queue. Investigators look only at content flagged by AI and make a publishing decision. However, while the decision feeds back into the algorithm, the reasons for the decision that the investigator took do not, making the current process very linear and literal. More dynamic algorithms are needed to keep pace with evolving bad actor tactics.

A next generation of AI tools will be able to identify and score a much larger range of attributes than just objects within the content itself. The content source (such as various attributes of the publisher) and context (such as geography, time of day, or relevant social, political or market events) both carry a relative risk factor that the content is illicit, illegal or inappropriate. In the not too distant future, algorithms will incorporate a huge multivariate set of content and context attributes. Based on attribute scoring a relative risk score will be calculated that will determine if something should be posted immediately, posted but still reviewed, reviewed before posting, or not posted at all. Attributes will be tracked over time and the feedback loop to track bad actor activity will become more accurate and nearly instantaneous.

Yet, even as AI becomes hyper accurate and more contextual at assessing content, the need for investigators will remain. Investigators bring the subject matter knowledge to make decisions that lie in the complex gray areas of decision-making. They bring empathy and contextual understanding that are important to content assessment. For example, they can view content from the subjective perspective of the content creator. They can also bring an understanding of cultural context to assess how content will be perceived by those in different demographics and geographies. But, unlike today, this new investigator will be highly trained and laser-focused on investigating complex, difficult content forensics.

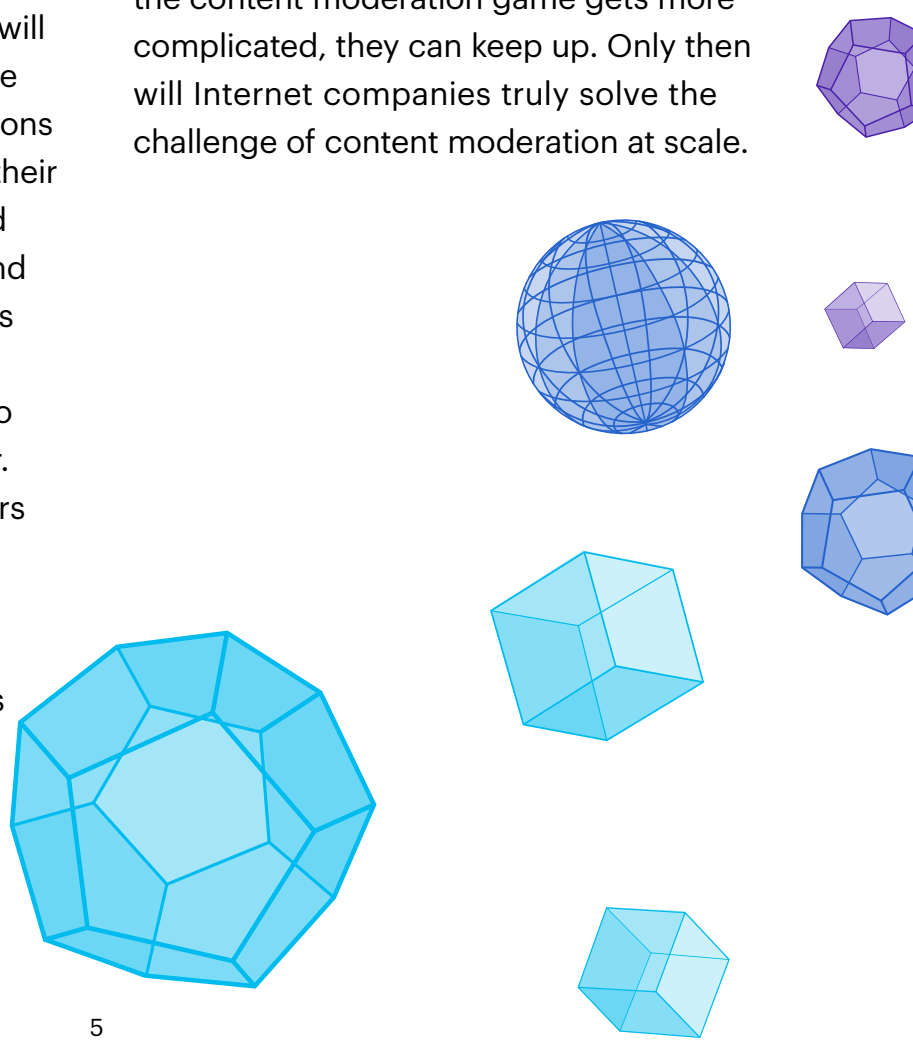
Armed with dynamic algorithms, advanced analytics and an amalgamation of data such as historical activity, account profiles, and the volumes of personal and behavioral data resident in most platform companies, these forensic investigators can develop a deep understanding of "bad actor" personas and begin to detect and even predict bad behavior at the source. For example, as AI becomes more sophisticated at detecting fraudulent content, fraudsters also become more sophisticated at evading detection. Once AI can provide the investigator with

intelligence on the content producer versus flagging the content itself, then the investigator can evaluate the producer (such as validating and risk scoring the creator of the app, digital ad, emails or other content) and take action (such as banning the profile) that will protect against the fraud before it occurs. When both the AI and forensic investigator work together, investigators can attack the "root cause" of all kinds of problems before they occur, dramatically reducing the need to actually moderate digital content at all.

WHAT THE FUTURE HOLDS

Efficient and effective content moderation at scale will be defined by a more synergistic relationship between humans and AI. AI will focus on its strength – evaluating massive amounts of data across multiple dimensions in near-real time. Humans will focus on their strength – reading between the lines and understanding the cultural context around the content. Today's content adjudicators will be supplanted by investigators with analytical thinking and techniques akin to an actuary or financial crime investigator. This evolving role will require investigators to develop specialized skills in product, market, legal and regulatory domains. It will also require native-level cultural understanding and training in forensics and psychology. The ability to attract and retain talent in these functions will become a critical source of competitive advantage. So will access to the latest AI technologies. We predict that content investigation will become a respected and rewarding career path and considered a strategic role in the organization.

AI will indeed turn content moderation on its head—not by eliminating the role but by turbocharging it. Certainly the total number of people working in content moderation can be reduced, but only if Internet companies stay on the forefront of AI technology. Companies providing content moderation services will need to maintain an investment in predictive analytics, workflow AI and more so that, as the content moderation game gets more complicated, they can keep up. Only then will Internet companies truly solve the challenge of content moderation at scale.



TO LEARN MORE ABOUT HOW TO COST-EFFECTIVELY SCALE YOUR CONTENT MODERATION CAPABILITY CONTACT:

Kevin J. Collins
Managing Director
kevin.j.collins@accenture.com
Mobile: 1.650.303.4633

Saurabh Mohanty
Managing Director
saurabh.mohanty@accenture.com
Mobile: 1.408.857.1304

ABOUT ACCENTURE

Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. Combining unmatched experience and specialized skills across more than 40 industries and all business functions—underpinned by the world's largest delivery network—Accenture works at the intersection of business and technology to help clients improve their performance and create sustainable value for their stakeholders. With approximately 401,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives. Visit us at www.accenture.com.