

A young woman with long, wavy brown hair is smiling broadly while looking at her smartphone. She is wearing a blue and white plaid button-down shirt. The background is a light-colored wall with a grid of dark lines, possibly a window or a decorative wall. The lighting is warm and soft, suggesting an indoor setting with natural light.

Say What? Give Customers Facts, Not Confusion

Fight back against customer confusion with content moderation strategies

Overview

Customers form opinions about brands when they read a news story or online review, hear about a friend's experience, or visit the brand's website. Simply put, brand perception matters.

So how can companies monitor and foster positive customer perceptions? It begins with content moderation. The staid practice of reviewing and eliminating objectionable content has become more important than ever. With the rise in fake news sites and anonymous posts on public sites, companies are reassessing their communication strategies to avoid negative brand associations.

In addition, the fast-paced nature of social media and programmatic advertising means that brands have less control over where their messages appear or even how the company's brand is perceived. The stakes are high for brands to communicate the right message on content-driven websites often open to users.

This paper examines the challenges moderators face today in making sure that user-generated content and other messages meet the company's standards for acceptable content. It also explains why an effective content moderation strategy needs both human moderators and technology solutions.

The stakes are high for brands to communicate the right message on content-driven websites often open to users.

TABLE OF CONTENTS

- The Facts 3**
- What Does a Content Moderator Do? 5**
- Balancing Automation with Human Insight 6**
- Case in Point 7**
- How to Measure Results 8**
- Choosing a Partner 9**

The Facts

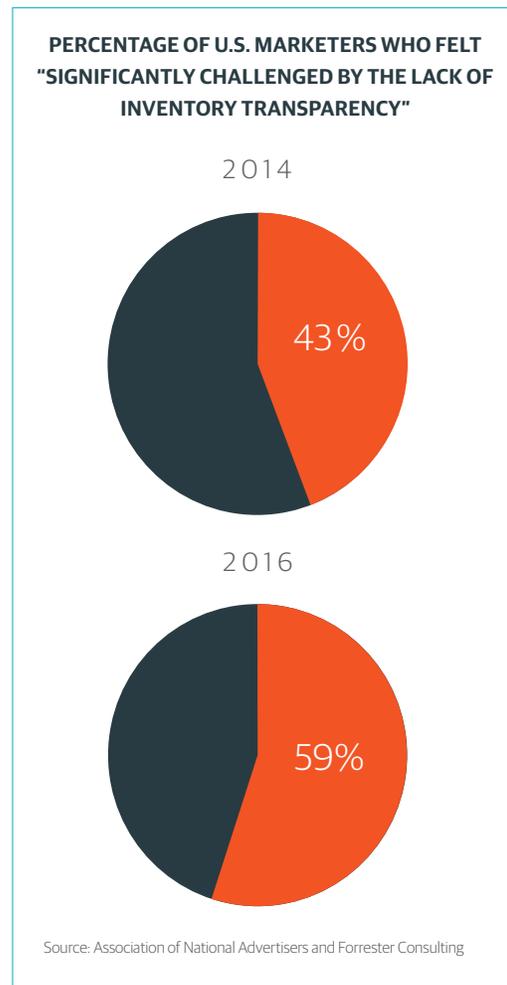
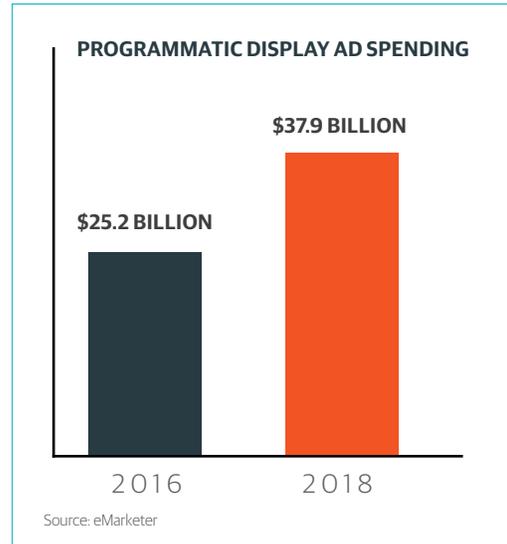
“Brand safety” has historically involved making sure a brand’s ads didn’t appear on inappropriate or disreputable websites. However, the rise of programmatic media buying—where ads are distributed across thousands of websites within milliseconds—complicates the moderator’s task. It can be difficult even for advertisers to know where their ads may appear.

In digital advertising, brands typically buy audience impressions or ads that are shown to specific groups of people. For example, a consumer may be tagged under “fashion” or “beauty” and will see these types of ads regardless of the site he or she is viewing. Advertisers buy based on these tags, and the ads are placed on sites that cater to certain customer segments. There is little “inventory transparency” about which sites the ads are actually placed on.

These are some of the factors that caused brands to suddenly find themselves mired in controversy. Kellogg’s, for example, was hit with a social media storm when the company’s ads unknowingly appeared on a website known for anti-Semitic, sexist, and racist articles. Kellogg’s announced that it was pulling all advertising from the website on the grounds that the site did not align with its values.

At the same time, programmatic media buying continues to grow. U.S. programmatic display ad spending is expected to reach \$37.9 billion, or 82 percent of total spending on display ads in 2018, up from \$25.2 billion in 2016, according to eMarketer.

Advertisers expressed fears about inventory transparency even before the fake news and hate speech phenomenon gained attention. According to a survey from the Association of National Advertisers (ANA) and Forrester Consulting, 59 percent of U.S. marketers who responded said they felt “significantly challenged by the lack of inventory transparency” in 2016, up from 43 percent in 2014.





Tools such as whitelists, blacklists, and semantic technologies are designed to help advertisers filter out objectionable sites and placements. Whitelists allow advertisers to select the sites where they want their ads to appear, while blacklists allow advertisers to do the opposite—indicate the sites on which they don't their ads to appear. Semantic technologies allow advertisers to prevent their ads from appearing on certain sites or next to certain types of content by filtering for language.

But research suggests that few marketers are actively using these safeguarding tools. Only 51 percent of U.S. marketers aggressively update blacklists, while 45 percent use whitelists, according to research from ANA and Forrester Consulting.

Even when these tools are used, they are not foolproof. Advertisers can't just launch an automated solution and forget about it. They must regularly check their ad placements and the sites they appear on to make sure nothing slips through the cracks. Facebook's algorithm, for instance, mistakenly blocked bsdetector.tech, the website for a popular browser extension that detects and tags fake news sources.

To tackle these messaging challenges, companies need a comprehensive strategy that enables them to proactively weed out problematic content quickly and efficiently. Enter content moderation.

Tools such as whitelists, blacklists, and semantic technologies are designed to help advertisers filter out objectionable sites and placements.

ONLY

51%

of U.S. marketers aggressively update blacklists

45%

of U.S. marketers use whitelists

Source: Association of National Advertisers and Forrester Consulting

What Does a Content Moderator Do?

These days, nearly every business has a social presence, whether it's publishers, consumer brands, or even B2B enterprises like Oracle or Cisco who have their own accounts on Twitter, Instagram, and Facebook. Content moderators play a critical role in monitoring user-generated content on these social sites and communities, ensuring that it fits within a brand's guidelines of acceptable content. A content moderator may handle any of the following jobs:



Pre-moderation

When users submit content to a brand's social account or community forum, the company may choose to have a moderator review it before making it visible to all members. This approach is best suited for content that is not time sensitive and/or includes legal risks where a moderator is needed to block libelous content. It could be helpful for example, in online communities targeted at children, as a way to prevent bullying and other inappropriate behavior.



Post-moderation

This role differs from pre-moderation because the user-generated content appears instantly online, while at the same time, it is placed on a virtual queue for moderation. This allows users to engage in real-time discussions without impeding the flow of conversation. Objectionable content on a Facebook ad or post, for instance, is flagged for the assigned team to review and potentially remove quickly.



Reactive moderation

Reactive moderators rely on users to report inappropriate content when they see it. This means posts are usually only checked if a complaint has been made, such as via a "report this" button. When alerted, an internal team of moderators will review the post and remove it if necessary. Reactive moderation puts power in the hands of the user. It shares the responsibility for blocking objectionable content with members. On the other hand, reactive moderation can result in a lot of false positives, where users have flagged content for no good reason.



User-only moderation

In this approach, users are generally responsible for deciding how useful or appropriate the user-generated content is. If a post has been flagged a certain number of times it will be automatically hidden. And similar to reactive moderation, this type of moderation can be easily scaled as the online community grows, with minimal costs. However, trusting a community to self-moderate itself carries obvious risks. Therefore, a company should consider appointing staff members to review the flagged content and/or use an average score to determine whether the content should remain visible or be reviewed.



Automated moderation

Automated moderation consists of deploying various technical tools to process user-generated content and apply defined rules to reject or approve submissions. The most commonly used tool is the word filter, in which a list of banned words is entered and the tool either replaces it with alternative word or blocks it. A similar tool is the IP ban list. There are also a number of more recent and sophisticated tools being developed. These include engines with automated conversational pattern analytics. Of course, the drawback is an automated solution could miss nuanced phrases or other irregularities that a human content moderator would catch.



Balancing Automation with Human Insight

To be effective, human content moderators must be well-acquainted with the site's content and its audience's tastes, in addition to being knowledgeable about the topics that are discussed on the site. They must also be steeped in the relevant laws governing the site's country of origin and be experts in the user guidelines and other platform-level specifics concerning what is or what is not allowed. And most important, they must be able to make quick decisions.

It is impossible, though, for even the fastest moderator to keep up with the constant stream of content that floods the Internet. Increasingly, companies are using a combination of technology and human intervention to flag objectionable content.

Facebook, for example, uses algorithms and human moderators to help it flag fake news. Algorithms tally fake news signals and prioritize what's sent to the fact checkers. If an article is flagged a certain number of times, it is directed to a coalition of fact-checking organizations including PolitiFact, Snopes, and ABC News. Those groups will review the article and offer recommendations on whether it should be marked as a "disputed" piece, a designation that will be visible to Facebook users.

Additionally, disputed articles will appear lower in the News Feed. If users still decide to share a disputed article, they will receive a pop-up message reminding them that the accuracy of the piece is questionable.

Increasingly, companies are using a combination of technology and human intervention to flag objectionable content.

Case in point:

Wellness company creates social interaction playbook

A wellness company was growing its social media presence and saw an increase in customer support requests through this channel. They were not adequately staffed internally to handle this volume, however, nor did they have a strategy for engaging customers, meeting needs, and representing the brand through social media.

We helped them design a full social media model and playbook to help engage with customers appropriately, and ensure a consistent brand and customer experience. The playbook documented processes and rules of engagement for the social media specialists. It illustrated resolution paths for various customer needs and a tiered support structure that defined needs that should be addressed by the specialists. Processes included content moderation guidelines, such as how to distinguish between appropriate public versus private discussions.

The team also implemented a quality assurance strategy to measure the specialists' performance, as well as a system for tracking volume, sentiment, and topic trends that emerged from the interactions.

THE RESULTS

90%

of the posts were answered within the allotted response times (30 minutes for Twitter, 60 minutes for Facebook, 60 minutes for other channels)

20%

reduction in staffing costs after we proved overnight staffing was unnecessary

92%

Average quality assurance performance that exceeded expectations





How to Measure Results

Measuring the effectiveness of content moderation requires more than adding up the number of blocked posts. It's possible, for example, that only a few people had published objectionable posts that day, thereby skewing the results.

A more effective approach is to examine a range of factors. Response times, as well as comparing the number and types of objectionable content that is being flagged are

important factors. Furthermore, what can the company learn about the content that is being removed or blocked? Is there a system in place to update a knowledge management database with these learnings?

Ultimately, the performance metrics of the content moderation team as a whole must be determined on a case-by-case basis depending on the company's objectives and goals.



Choosing a Partner

A successful content moderation strategy needs both human moderators and technology solutions. But not every company has the expertise or resources necessary for building an effective content moderation team. Working with an experienced partner to build a team offers a number of benefits.

A trusted partner who acts as an extension of the client's brand can help companies represent their best selves online. The right partner has the ability to recruit and train moderators who understand the brand's culture and goals and provides round-the-clock support.

Look for a customer experience partner who will work with you on designing a content moderation strategy that fits your company's specific needs. Furthermore, the partner's hiring practices should be customized to each position and allows the company to scale quickly and efficiently.

For example, an automated solution that uses text analytics and machine learning can flag questionable content that is brought to a contact center associate's attention. The associate, who has been trained on the appropriate response, will review the content and determine whether or not to remove it. Additionally, agents could moderate content during a lull in answering calls or emails, to further optimize their time.

Conclusion

Content moderation is at a turning point. The rapid and constant nature of digital media requires companies to take a more active role in monitoring content that affects their brand. But this is not something marketers can set and forget. Effective content moderation strategies need technology to quickly comb the Internet for problematic content as well as well-trained associates to separate the false positives from the actual problems. Brands that are prepared to handle emerging challenges are the ones that win positive mindshare and customer loyalty.

About TTEC

TTEC (NASDAQ: TTEC) is a leading global provider of customer experience, engagement, growth and trust and safety solutions delivered through its proprietary end-to-end Humanify™ Customer Engagement as a Service offering. Founded in 1982, the Company helps its clients acquire, retain, and grow profitable customer relationships. Using customer-centric strategy, technology, processes and operations, TTEC partners with business leadership across marketing, sales and customer care to design and deliver a simple, more human customer experience across every interaction channel. TTEC's 49,500 employees live by a set of customer-focused values that guide relationships with clients, their customers, and each other. To learn more about how TTEC is bringing humanity to the customer experience, visit ttec.com.



solutions@ttec.com
+1.303.397.8100 (outside the U.S.)
ttec.com