Stopping Content Abuse Before It Happens

sift science

Contents

The power of user-generated content 1

The growing problem of content abuse 3

Why does content abuse matter? 5

Reactive solutions: More trouble than they're worth 7

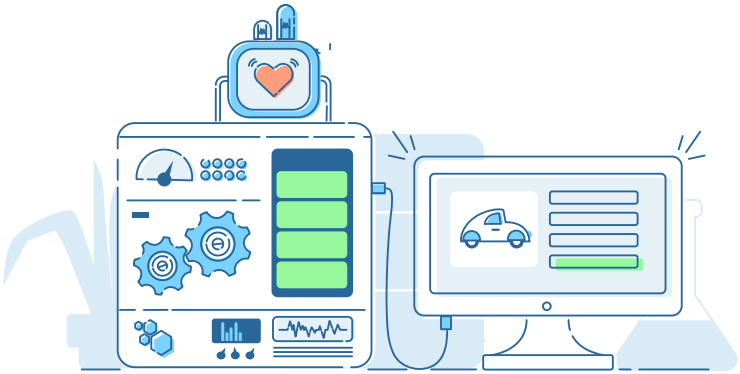
A challenge and an opportunity 8

What would you do if you could trust your users? 10

How does Sift Science make a difference? 12

The power of user-generated content

Think of the most successful online businesses. What do they have in common? An overwhelming portion of the most influential online marketplaces and communities are powered by user-generated content (UGC).



UGC can include comments, blog posts, videos, marketplace listings... anything that a user might contribute to a website. It's easy to see why it's so powerful. Many websites' business models are designed around UGC, which makes high-quality UGC a core part of the customer experience.

UGC is the lifeblood of online marketplaces and communities like Twitter, Pinterest, YouTube, Airbnb, and Indeed.

If they can't attract people to contribute to their site, then they don't have a business. Their future is in the hands of their users.

UGC success story

Remember the Ice Bucket Challenge? The Ice Bucket Challenge generated millions of donations – about $100 million – far more efficiently than any previous donation campaign. *The secret?* **UGC**.

In just one minute...

people post
455,000 Tweets

users watch
4,146,600 videos

Instagram users upload
46,740 million posts

people post **510,000**
Facebook comments

social media gains
840 new users

Source

5 Ways UGC Drives eCommerce

1. UGC is 35% more memorable than other media

2. Customers spend 5.4 hours per day with UGC

3. User-generated YouTube videos get 10x more views than those uploaded by brands

4. UGC results in 29% higher web conversions than sites or ads without it

5. 25% of search results for the world's 20 largest brands are links to UGC

The growing problem of content abuse

But content abuse can have an equally powerful, albeit negative, impact on your bottom line. Content abuse occurs when fake or malicious UGC is created or shared by scammers, to defraud the business or another user. Any company that considers content to be a core part of their customer experience is at risk for content abuse.

Content abuse impacts both your top line and bottom line. It hits your user base that you've worked so hard to acquire, delight, and retain. Good users who experience (or even see) abusive content will feel reluctant to engage with the community and are likely to churn. They'll tell their friends about it, too.

This is different from traditional payment fraud, in which a user might pay with a stolen credit card, and you're responsible for the chargebacks. With content abuse, your company is the victim and you lose money...but it's much harder to calculate the true cost.

Unlike credit card fraud or chargebacks, which are clearly defined problems, content abuse can feel like a moving target. It's difficult to fight, in part because it can take myriad forms..

The many faces of content abuse

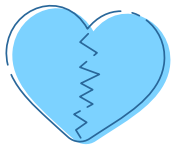
Spam

Most internet users are all too familiar with spam. It's any unsolicited advertising or other message that's usually sent to a large number of users. While it used to be primarily sent via email, anti-spam filters have gotten so effective that spammers have turned to social sites and messaging apps to get their message across. Whether it's a promotion for someone's website posted in a comment, a notification prompting you to "like" a phony page, or a malicious link hidden in a chat message – spam shows no sign of going away.

Scams: Fake listings

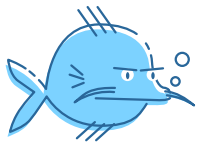
Scammers sometimes post fraudulent listings or counterfeit goods on online marketplaces, promising goods or services that they have no intention to deliver. The listing may be counterfeit, or it may not even exist.

When an unsuspecting user tries to transact with the fraudster, the fraudster might trick them into giving up their personal information, or paying for the good or service off-platform, where they are not protected by the marketplace. When buyers on two sided marketplaces purchase a good and don't get what they paid for, they are less likely to return to that marketplace. When they do return, they're more likely to be skeptical of listings, and therefore get less value from the site.



Phishing

Fraudsters often pose as legitimate users to trick their victims into giving up personal information, such as their bank account info or credit card number. For instance, by posting fake job listings, scammers can gain a wealth of personal information from unsuspecting applicants.



Catfishing

A big headache on dating sites, catfishing occurs when a scammer impersonates someone to gain a victim's trust. For example, on dating sites, many fraudsters pretend to be an attractive figure like a soldier so that their victim will engage with them. They then scam the victim of their credit card info, login credentials, etc. In 2016, imposter scam complaints surpassed identity theft as the most common type of consumer complaint. The Federal Trade Commission received 400,000 complaints that year – and those are just reported instances of imposter scams.

Fake reviews

Consumers rely on reviews to make spending decisions. As of 2016, Yelp had about 121 million reviews; Goodreads had 50 million. Bad actors use fake and malicious reviews for phishing, malware distribution, spam, and other harmful ends. These users often haven't even used the service, but that doesn't stop them from making an impact.

Toxicity

Sites that rely on UGC will inevitably attract rabble-rousers. Hate speech, profanity, or other inappropriate content can be detrimental to a business's brand. Ultimately, users want to trust that they can engage with a website without being subjected to uncomfortable content. If a business can't make that promise, then potential users will turn elsewhere.

Why does content abuse matter?

Content abuse degrades a marketplace or community, having a detrimental impact on your brand and bottom line. It's easy for it to snowball out of control: When scammers start posting fraudulent content on a platform, good users leave, and bad users flock to the site. A marketplace or community can quickly form a negative reputation. If the site relies on UGC, word of mouth can spread fast, amplifying the voices of those who are disparaging the site.

Here's how content abuse can harm your business:

You can lose customers

Content abuse sends a message: your business isn't trustworthy. If customers hear that a website is a magnet for content abuse, they may worry about their own safety and stay away from the business...and tell their friends to do the same.

User engagement can decrease

Businesses that rely on UGC need consistent engagement to stay afloat. If the business is an online community, that usually means posts and comments; if it's a marketplace, that typically means items for sale. Without these drivers, your bottom line will suffer. If content abuse runs rampant, people won't engage with the site.

Your site can decline in search rankings

If your business continues to suffer from a lack of engagement, your search ranking will take a hit, too. That further decreases the chances that a potential customer will find your site.

You can be financially and legally liable for damages

Let's say a fraudster scams a user on your marketplace out of their credit card number, or someone posts hate speech targeting someone in your online community. Depending on the circumstances and severity of the damages, the business could be forced to pay up in court. That means legal fees – and usually bad press.

Your customer acquisition cost can increase

If your site is perceived as untrustworthy and people aren't engaging, then each new customer becomes an expensive and hard-won investment.

Your brand and bottom line can take a hit

If left unchecked, content abuse can lead to bad reviews, negative press, and a drop in customers.

Exposure rate: how many people are seeing bad content?

As a marketplace or community with user-generated content, engagement metrics are all-important. But fraud can heavily affect daily and monthly active users – which means it pays to get ahead of the curve and prevent fraud before it impacts your platform and community.

One metric we use to quantify the damage done by fake content is "exposure rate." This calculates how many people across your ecosystem are seeing the content – which is more important than how many individual pieces of bad content are being created. Think of it this way: a single piece of content viewed 1,000 times = 1,000 pieces of unique content viewed 1,000 times.

To calculate the exposure rate for a single piece of content:

Based on user reports

Take the number of customer alerts (calls, flags, emails) about that content and multiply it by the number of total impressions of that content. Then, divide the result by the total impressions on your site. You'll have your overall exposure rate.

Keep in mind: *Not everyone reports bad content, so this estimate may be on the lower end. Also, customers may incorrectly report content that isn't actually spam.*

Based on a sample

Take a sample of listings or posts and identify which ones are fake or spammy. Take the number of spammy or fake content pieces you find and divide by the total number you sampled. Multiply that number by total the total impressions on your site.

Keep in mind: *You'll need to sample a large enough number of content pieces to derive a meaningful result.*

The power of trust

Websites that rely on user-generated content have an implicit contract with their users. Customers trust businesses to provide a service that's secure and reliable. Your users depend on you to keep their community and activities safe from fraudsters. In return, businesses trust customers to participate in a community or marketplace in good faith. When content abuse runs rampant, it's not just bad business: it's a breach of trust.

More than ever, fraudulent content is eroding the digital trust between users and businesses. But how do you fight fraud when you can't tell who the fraudsters are? How can you build an online community if you don't know who to trust?

Reactive solutions: More trouble than they're worth

Fighting content abuse isn't as simple as deleting a bad piece of content. Once you've identified content abuse, you have to decide whether this is a one-off instance from an otherwise honest member of the community, or whether this incident warrants a ban. These are **crucial decisions**. A business that accidentally blocks users for something that isn't content abuse risks driving away good customers. And a business that doesn't do enough to ban fraudsters will alienate honest users.

The other problem? Content moderation is **cumbersome and slow**. By the time your team has detected fraudulent content, users on your site have already seen it. The damage has already been done. Your moderation team can only clean up the fallout.

Despite these issues, most businesses currently use **reactive solutions** like content moderation teams. These teams are usually in-house contractors who maintain blacklists and delete fraudulent content. Their focus is cleaning up the mess after the fraud has occurred rather than working to preempt the fraud.

Many fraud teams use rules-based systems to combat fake content. The team creates a series of rules to parse fraudsters from honest users, and when the rules are broken, the system bans the offending user. Rules-based systems require **large, unwieldy teams** to supervise and update. To make matters worse, rules quickly become outdated as fraudsters evolve. Large fraud teams, especially those with international contractors, often find it difficult to communicate rules changes across the organization, or keep up with new trends and languages. These lapses in communication result in **reviewer bias and false positives**, exacerbating an already complex problem.

> Content moderation is cumbersome and slow. By the time your team has detected fraudulent content, users on your site have already seen it.

A challenge and an opportunity

Content abuse isn't going away. In fact, it's more deceptive and pervasive than ever. Fraudsters are now using artificial intelligence systems to create realistic fake content. Content abuse has become so sophisticated that by 2022, Gartner predicts that people will be exposed to false information more often than they'll see true information. In other words, people will begin to approach user-generated content with the expectation that it might be – or is – fake.

This dramatic shift will fundamentally change the way we interact online. Content abuse is creating a world in which responsible users cannot and do not trust the information they consume on the internet. **Consumers *expect* to encounter fraud**. Their experiences of online communities are marred by their suspicion of other users, who might not be who they say they are. Buyers and sellers on online marketplaces may be reluctant to list their items, or hesitant to make purchases. Users might wonder why they should bother contributing to a community comprised of fraudsters.

Content abuse poses both **a challenge and an opportunity** for businesses. Businesses are under increasing pressure to detect and stop fraud before it hurts their customers. But they're floundering.

> Content abuse has become so sophisticated that by 2022, people will see false information online more often than they'll see true information. – Gartner

Businesses that can earn their customers' trust will stand out in this landscape. And since most companies have yet to implement a robust system to fight content abuse, the first to do so will leap ahead of the competition.

Proactive solutions: Stopping content abuse before it happens

Reactive solutions aren't enough to stop content abuse. Content moderation alone won't enable you to stay ahead of the game. Rules-based systems are not scalable. An effective fraud solution must optimize across speed, accuracy, and intelligence. Businesses that invest in manual solutions risk indelible damage to their brand and bottom line.

UGC and Millennials: 5 Stats to Know

1. 59% of Millennials use UGC to inform purchasing decisions
2. Millennials trust UGC 50% more than brand-generated content
3. 71% of Millennials engage with UGC daily
4. Millennials find UGC 35% more memorable than other forms of content
5. 84% of Millennials say UGC on company sites influences what they buy

Only machine learning (ML) systems are up to the task. Reactive solutions like rules-based systems plod along behind speedy fraudsters, but **ML solutions get smarter and faster** each time fraud occurs on your system. In stark contrast to reactive fraud solutions, **ML solutions are proactive**, stopping fraud before it happens. Fraudsters use advanced systems and methods to exploit your business. Rules can't adapt fast enough. You have to fight fire with fire.

Some of the most powerful companies and organizations in the world are turning to ML to fight content abuse. Last year, Facebook invested heavily in ML solutions to cut down on content abuse. Google did the same. Inundated by fake content on Google Maps and Google My Business, the company developed ML systems that decreased fake listings by 75%. Google can now detect 85% of fake listings before they even make it to Maps. In 2016, a cybersecurity team at Cisco developed an algorithm designed to combat fake reporting. Even The New York Times relies on ML to weed out fake and abusive comments on news articles.

What would you do if you could trust your users?

Content abuse, including scams, spam and other malicious user-generated content, is one of the fastest growing types of fraud. As commerce, content, and users move online, protecting your community from malicious attacks is becoming more critical than ever before.

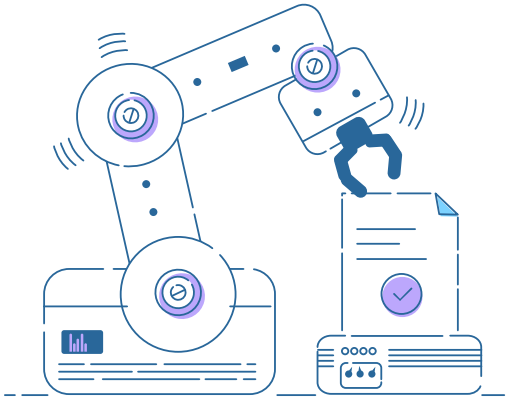
Sift Science uses Live Machine Learning™ to help you identify and stop bad users before they can post content. Our Content Abuse Prevention solution protects the community from multiple forms of content abuse, helping leading businesses across the world automate content review, improve brand perception and conversion rates, and increase user engagement and retention rates.

From bulk attacks to sleeper accounts, we've helped many customers uncover fraudulent content and behavior that was previously impossible to discover on their platform, optimizing user experiences for their good customers.

Catch more abusive users, faster, without increasing headcount, and see the volume of flagged content go down by up to 70%.

What sets Sift Science apart? Existing rules-based or manually-intensive solutions are reactive, expensive, and difficult to maintain as your business grows.

The Sift Science solution includes built-in automation tools and easy-to-use dashboards to help you stop content abuse in real time.



We use 16,000+ signals to get the most holistic view possible. Our transparent scoring makes it easy to understand why a user is risky or safe. Our machine learning models, including deep learning and natural language processing, unlock insights from each piece of content and data point.

Content Analysis + Behavior Footprint

Our unified model analyzes the content as well as how your users create it, picking up on signals like timing and sequence of behaviors, velocity of different activities, and unique users they interact with.

Global and Custom Models Across Our Network

Sift Science tracks fraudulent behavior across our global network in real time. Any time fraud is attempted on any of our customers' sites or apps, our model instantly learns from and prevents similar behavior on your site.

Language Agnostic Algorithms

Our models automatically pick up fraudsters' evoliving language patterns, so your team doesn't have to constantly monitor, identify, and blacklist new terms.

Enterprise Security and Scale

We protect millions of user accounts every day, ingesting over 3,000 events per second, with 200ms score latency and 99.95% uptime. We are SOC 2 compliant, and our data protection policies align with GDPR and Privacy Shield.

How does Sift Science make a difference?

Stop fraudsters early

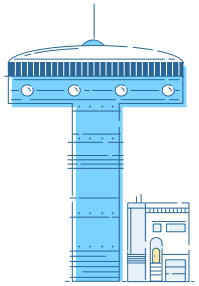
Don't wait for fraudsters to act. Examine bad users' behavior in real time and stop them before they create their fake listings, spammy comments, and malicious messages.

Catch more bad content, faster

Identify and remove malicious content before the community sees or flags it, drastically reducing its exposure rate.

Automate away your risks and worries as you scale.

Protect your community without worrying about the need for larger content moderation teams.



Grow your business with trust

Build better communities.
Grow your top-line revenue by improving the user experience for your good customers.

Catch more fraudsters and cut flagged content by 70%...without increasing your headcount.

Customers and Results

Find more bad users than you are finding today
We helped a major dating site find 35% more bad users than their existing system, and brought another dating site's false positive rate to nearly zero.

Increase speed and accuracy, all with your existing headcount:
We helped one of the largest social marketplaces for fashion detect 70% of scam and spam content hours before the community flagged it.

"Sift Science enabled us to make the user buying experience more friendly and frictionless allowing us to provide a better user experience."

"Fraud management at SeatGeek is built upon the Sift Science solution, with payment abuse and content abuse findings providing insights into how trustworthy a buyer or seller is."

"There is no way we could do what we do without Sift Science."

"We no longer simply react to fraud but can now take a proactive approach to prevent it and create better efficiencies for our business."

"Patreon trusts Sift Science. If Sift Science gives us a score of 90, we know that that's fraud – there's no need to double check. Our error rate is as close to zero as it can get."

The Sift Science Digital Trust Platform

One scalable solution for every vector of abuse

Content Abuse Prevention is just one part of the Sift Science Digital Trust Platform. With a single integration, you have access to a suite of products running on a single platform to battle every type of fraud and abuse, including account takeover, payment fraud, promo abuse, and more.

Account Abuse

Promo Abuse

Account Takeover

Payment Fraud

Content Abuse

For more information, visit https://siftscience.com/products/content-abuse or contact sales@siftscience.com.