The Nefarious
User Generated Content Report

WebPurify's Report on The Threat of Unmoderated User Generated Content

WebPurify

User generated content (UGC) offers the incredible opportunity to deepen a brand's connection with its online customers through enhanced personalization. However, lurking within all the friendly and helpful submissions are the bad apples, exposing users to illicit or disturbing content notwithstanding a brand's clear guidelines. As UGC becomes further woven into the online experience, the threat to all internet users – but especially children – will continue to grow. Since 80% of Americans say UGC highly impacts their purchasing decisions, brands must create a safe environment for users or risk losing business.

Our 'Nefarious User Generated Content Report' reveals that without the right moderation safeguards in place, companies across all sectors are at risk of exposing their online users to thousands of illicit images per day, as well as livestreamed traumatic events, which have seen a marked increase in the past decade. Meanwhile, as children explore the internet and enter virtual spaces at increasingly younger ages, they become even more vulnerable to exposure to inappropriate content and exploitation by predators. In another perhaps less disturbing but still worrisome example of content headaches, eCommerce companies face a growing exposure to liability when customers personalize products. The wrong image on a branded item can lead to copyright infringement, PR fire drills and damaged corporate reputation.

Companies are also starting to integrate generative AI tools into their user experience, opening up an additional method to create inappropriate text and images. Users are already attempting to submit offensive content across a wide variety of sectors, so we'll likely see the same in their submissions to generative AI tools. Even more alarming, the higher the proportion of inappropriate content online, the more likely AI models will generate illicit content pulled from it, even if the prompt is harmless. With AI producing its own images and text, we've created yet another path for users to potentially be exposed to harmful content.

User- and AI-generated content will become increasingly essential to maintain customer engagement, so brands must employ both AI tools and human content moderators to identify compromising content, to protect vulnerable customers and their user experience. With proper safeguards in place, companies can confidently enfranchise users to express their individuality, safely explore AI integrations, build connections to the brand, and have a positive overall experience, while mitigating the risks inherent in self-expression gone too far.

Jonathan Freger
CTO and Co-Founder
WebPurify

Overall user submissions

1 in 27

user submissions is flagged as not safe for work (**NSFW**)

1 in 19

user submissions is not safe for the brand (**NSFB**)

NSFW User Submissions by Sector

Dating sites:
1 in 9 submissions

Video games:
1 in 14 submissions

eCommerce:
1 in 35 submissions

eLearning / Education:
1 in 66 submissions

Telecom:
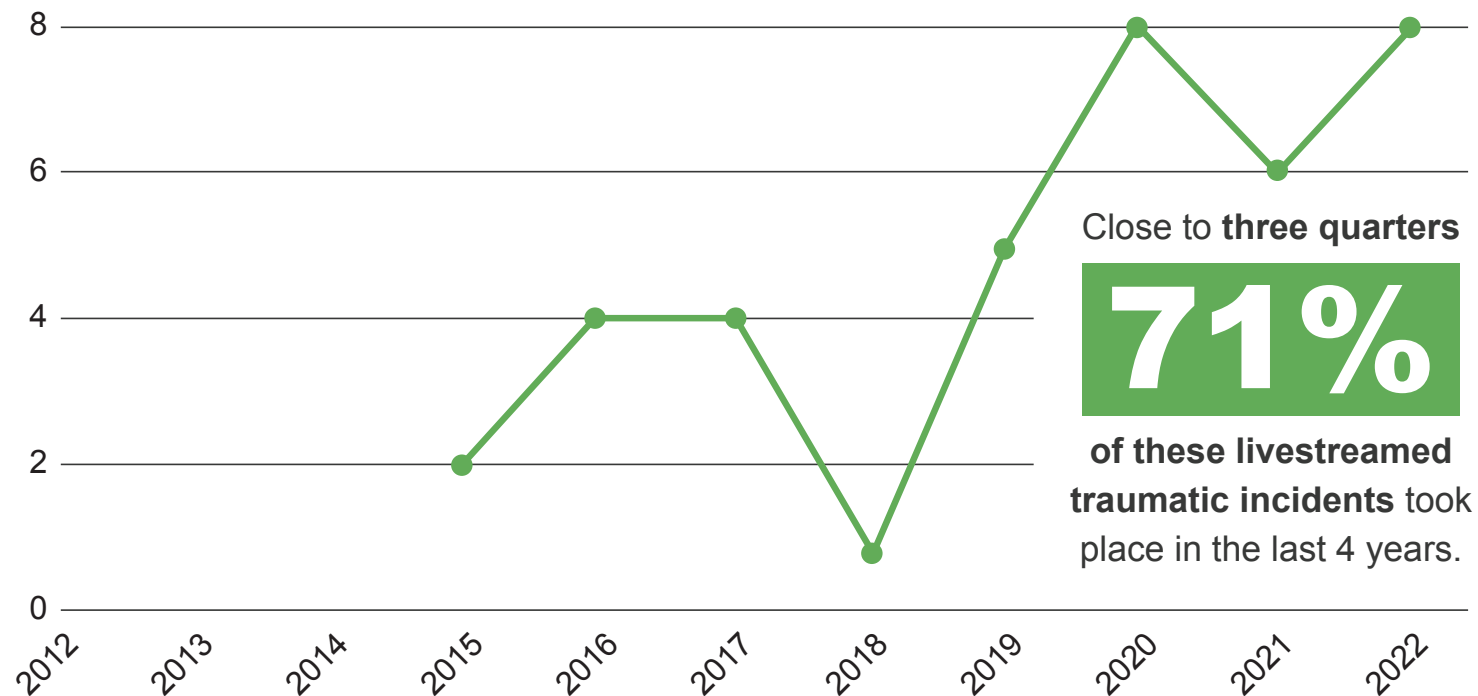1 in 76 submissions

Financial Services:
1 in 80 submissions

Food and Beverage:
1 in 89 submissions

Widely Shared Traumatic Livestreamed Events Are Proliferating

There have been roughly **38 broadly visible violent livestreamed event**s, such as the New Zealand Christchurch shootings or the US Capitol riots, over the past decade

Unfortunately, for a multitude of reasons, the number and frequency of these incidents is **accelerating** in recent years

Critical Livestreamed Events

Close to **three quarters**

71%

of these livestreamed traumatic incidents took place in the last 4 years.

8

6

4

2

0

2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022

Children Online

WebPurify alone contributed to the **arrest of over 200 child sexual predators in 2022**

Over **1 in 19** images WebPurify reviewed were considered appropriate for adults but threatening to children

eLearning sites can also harbor potentially damaging content. **1 in 66** submissions on these sites was deemed NSFW.

Many inappropriate submissions also contain content popular with children. WebPurify found that year over year, both **'SpongeBob SquarePants' and 'The Simpsons' top the list of TV shows that most often appear or are referenced in illicit user generated content.**

Personalized Product Submissions can Endanger eCommerce Brands

WebPurify's eCommerce clients received over **19 million personalized product submissions** in 2022
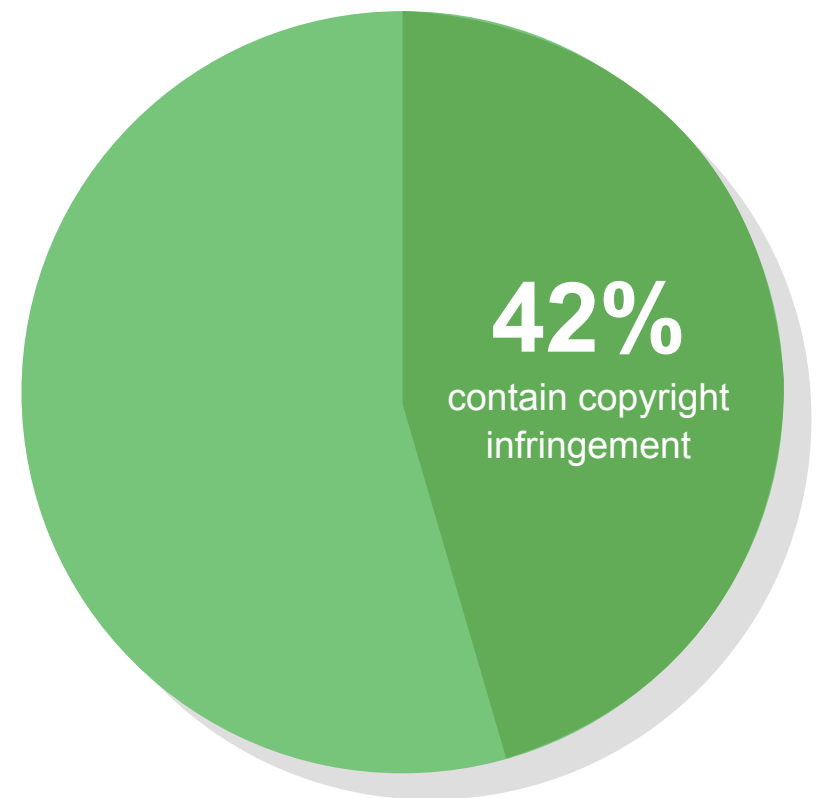
Daily Personalized Product Submissions
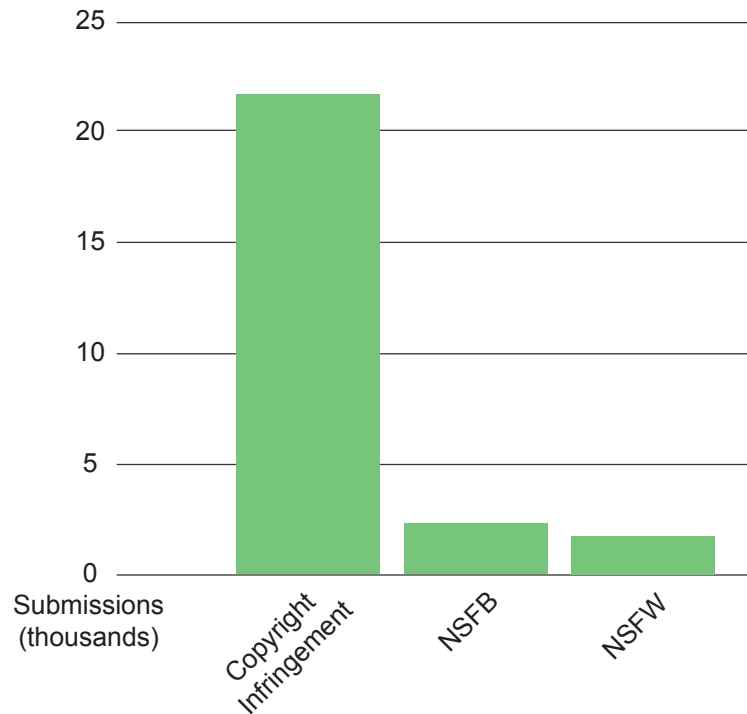
In 2022, WebPurify's eCommerce clients received on average:

- Over **21,700** submissions containing copyright infringements a day
- Over **2,700** NSFB submissions a day
- Over **2,000** NSFW submissions a day

Of all personalized product submissions in 2022

42% contain copyright infringement

Research Methodology

WebPurify analyzed its internal databases across all client categories including dating apps, product customization, eCommerce, eLearning/education, telecom, financial services, food and beverage, and social media to determine the instances of NSFW and NSFB user generated content. All data was pulled from January 1, 2022 – December 31, 2022, except for data on top TV shows that were featured in illicit UGC and the widely shared violent event stats, which used data from the past 4 and 10 years, respectively.

When designating submissions as "appropriate for adults but threatening to children," WebPurify followed the guidelines that govern its image moderation tools, grouping content labeled as 'Partial Nudity' and 'Sexually Suggestive' into this category.

For the purposes of this report, WebPurify classifies "widely shared violent events" as any crime that was live streamed and widely shared online. This does not include crimes recorded by police cams, bystanders, or CCTV and subsequently uploaded.

About WebPurify

The world's leading professional content moderation service, WebPurify combines the power of AI and human moderators to keep brands and online communities safe. With 17 years of experience (and counting) to its name, WebPurify is trusted by some of the world's most well-known organizations including Microsoft, PBS, Allstate, LegalZoom and beyond. WebPurify makes moderation accessible to businesses of all sizes, offering a range of solutions from plug-and-play to custom enterprise moderation. WebPurify is proud to in many respects function as the "first responders of the internet," protecting vulnerable groups from illicit content and supporting law enforcement in removing the worst offenders from our digital spheres. To learn more, visit: https://www.webpurify.com/.