YouTube still can’t stop child predators in its comments

YouTube is facing a new wave of criticism over the alarming number of predatory comments and videos targeting young children.

The latest concerns started with a Reddit post, submitted to r/Drama, and a YouTube video, exposing a “wormhole into a soft-core pedophilia ring on YouTube,” according to Matt Watson. Watson, a former YouTube creator who returned with a single video and live stream about the topic, demonstrated how a search for something like “bikini haul,” a subgenre of video where women show various bikinis they’ve purchased, can lead to disturbing and exploitative videos of children. The videos aren’t pornographic in nature, but the comment sections are full of people time stamping specific scenes that sexualize the child or children in the video. Comments about how beautiful young girls are also litter the comment section.

Although Watson’s video is gaining mainstream attention, this isn’t the first time that YouTube has dealt with this issue. In 2017, YouTube updated its policies to address an event known as “ElsaGate,” in which disturbing, sexualized kids’ content was being recommended to children. That same year, YouTube decided to close some comment sections on videos with children in an attempt to block predatory behavior from pedophiles. As early as 2013, Google changed its search algorithm to prevent exploitative content from appearing in searches on both Google and YouTube. But despite years of public outcry, YouTube still hasn’t found a way to effectively

deal with apparent predators on its platform.

The heart of the problem is YouTube's recommendation algorithm, a system that has been widely criticized in the past. It only took two clicks for Watson to venture away from a video of a woman showcasing bikinis she's purchased to a video of a young girl playing. Although the video is innocent, the comments below — which include timestamps calling out certain angles in the video and predatory responses to the images — certainly aren't.

"Youtube's recommended algorithm is facilitating pedophiles' ability to connect with each-other."
"Youtube's recommended algorithm is facilitating pedophiles' ability to connect with each-other, trade contact info, and link to actual child pornography in the comments," Watson wrote on Reddit. "I can consistently get access to it from vanilla, never-before-used YouTube accounts via innocuous videos in less than ten minutes, in sometimes less than five clicks."

The Verge has tried to recreate the situation multiple times, and through each experiment, discovered it took six clicks or less to find videos with predatory comments in the comment section. A statement from a YouTube spokesperson confirmed that several of the videos featured in Watson's video have since been removed.

"Any content — including comments — that endangers minors is abhorrent and we have clear policies prohibiting this on YouTube," the spokesperson said. "We enforce these policies aggressively, reporting it to the relevant authorities, removing it from our platform and terminating accounts. We continue to invest heavily in technology, teams and partnerships with charities to tackle this issue."

YouTube has tried a number of enforcement tactics, but none have prevented more scandals from surfacing. In November 2017,

Johanna Wright, the former vice president of product management at YouTube, issued a blog post announcing new ways the company would tackle the issue. Other than removing videos that were flagged or publicly reported, YouTube's most direct attempt at stopping the spread of these videos was to close down comment sections.

"Any content — including comments — that endangers minors is abhorrent and we have clear policies prohibiting this on YouTube."
"Comments of this nature are abhorrent and we work with NCMEC to report illegal behavior to law enforcement," Wright wrote in 2017. "Starting this week we will begin taking an even more aggressive stance by turning off all comments on videos of minors where we see these types of comments."

While YouTube claims that most of these videos now have comments turned off, that doesn't seem to be true in practice. Most of the videos *The Verge* came across had comments enabled. Many of the videos that appear in Watson's video also have comments enabled. These videos can be shared privately between users, and are full of timestamps or disturbing comments. YouTube uses a combination of machine learning technology and human reviewers, including those within the Trusted Flagger program, to discover these videos and remove them. With 450 hours of content uploaded every minute and billions of users logging in every month, some videos are bound to slip through.

Even for Trusted Flaggers, there are still roadblocks preventing quick action. "Content flagged by Trusted Flaggers is not automatically removed or subject to any differential policy treatment — the same standards apply for flags received from other users," one source, who asked to remain anonymous, told *The Verge*. "However, because of their high degree of accuracy, flags from Trusted Flaggers are prioritized for review. They have no say in what does or doesn't happen to reported content."

One of Watson’s biggest questions is why these videos are allowed to appear on YouTube at all. Most were uploaded by suspicious accounts. With the comment sections full of predatory comments, it shouldn’t have been hard to see that something was wrong. YouTube’s spokesperson declined to answer when *The Verge* presented this question. YouTube’s community guidelines do address videos that may be used to sexually exploit children, however, noting that “Uploading, commenting on or engaging in activity that sexualizes minors may result in the content being removed and the account being terminated.”

“They have no say in what does or doesn’t happen to reported content.”

While individual videos are removed, the problematic users are rarely banned, leaving them free to upload more videos in the future. When Watson reported his own links to child pornography, YouTube removed the videos, but the accounts that posted the videos usually remained active. YouTube did not respond to *The Verge*’s question about how the trust and safety team determined which accounts were allowed to remain active and which weren’t.

YouTube has clear rules surrounding content that contains children. One passage on its community guidelines and policies page states, “Sexually explicit content featuring minors and content that sexually exploits minors” is strictly forbidden, adding that content featuring these types of images will be reported to law enforcement. But many of the videos flagged by Watson aren’t violating YouTube’s guidelines, since videos that are being used by bad actors or inspiring disturbing comments aren’t specifically addressed by YouTube’s policies. Typically, the only action YouTube has taken is closing down comments on offending videos.

Increasingly, YouTubers like Watson are trying to do what YouTube isn’t: finding and calling attention to these types of videos. While most creators agree with Watson’s video, some

disagree with the tactics being used. Eion of popular YouTube channel, Nerd City, which often offers commentary and long-form discussions on a medley of YouTube topics, told *The Verge* that the video Watson made is one that he – or a number of other YouTube creators – could have produced for a long time. Eion chose not to for a number of reasons, including trying to help explain to his audience the balance between entertainment and illegal content on YouTube.

"A YouTuber and their viewers aren't authorized to analyze child porn," Eion said. "We're not allowed to download it, we're not allowed to view it, we're not authorized to determine what laws are being broken and what to do about it. That's not our role. You have to leave that to professionals. If you've actually discovered child porn, you don't need to create awareness of it. You need to report it."

"This danger is everywhere on the internet and volume is very big."
It's the same reason that other creators – including Daniel "Keemstar" Keem, who hosts the popular DramaAlert show – have stayed away from covering the story. He tweeted that although he agreed with Watson's opinions regarding the disturbing content on the site, he disagreed with the way Watson was going about the issue.

"YouTube is dedicated to remove child predators from the site," Keem tweeted. "They have banned millions of accounts, disabled comment sections and removed videos. This danger is everywhere on the internet and volume is very big. Trust me YouTube is not ignoring this issue."

YouTube has tried to tackle child sexual abuse imagery in the past by partnering closely with authorities and experts like NCMEC and the Internet Watch Foundation (IWF). These organizations provide YouTube's trust and safety team with digital fingerprints of reference files of known child sexual abuse content. The team then uses hashes to detect copies of

known illegal content and remove it from YouTube. The company also reports all child sexual abuse imagery and illegal comments, including messages, to proper law enforcement.

Exploitation videos make up only a fraction of a percent of the content that violates YouTube's guidelines, but the platform sees this specific area as a low-volume, high-risk area. The trust and safety team has said it's committed to fighting the issue by developing new tools and strategies — even when those measures sometimes go too far. Most recently, a number of *Pokémon GO* and *Club Penguin* YouTube channels were reportedly hit by wrongful terminations because of the term "CP" being used in the title, which YouTube's machine learning tools could read as "child pornography."

Most creators are aware that YouTube has a problem with this type of content — but they're also confident that YouTube is working on the issue. This isn't just an issue for the platform, but one for the internet at large, according to creators like Eion.

"I think this shows that people have a limited understanding of how search engines should work," Eion said. "You can find anything in Google by typing the search. You'll get shown relevant content to anything you type in. Now, if you find something doing that, and you think law enforcement needs to know about it, then report it."

"We're not law enforcement. I'm glad this issue is being talked about, but report it. Just report it."

https://www.theverge.com/2019/2/19/18229938/youtube-child-exploitation-recommendation-algorithm-predators