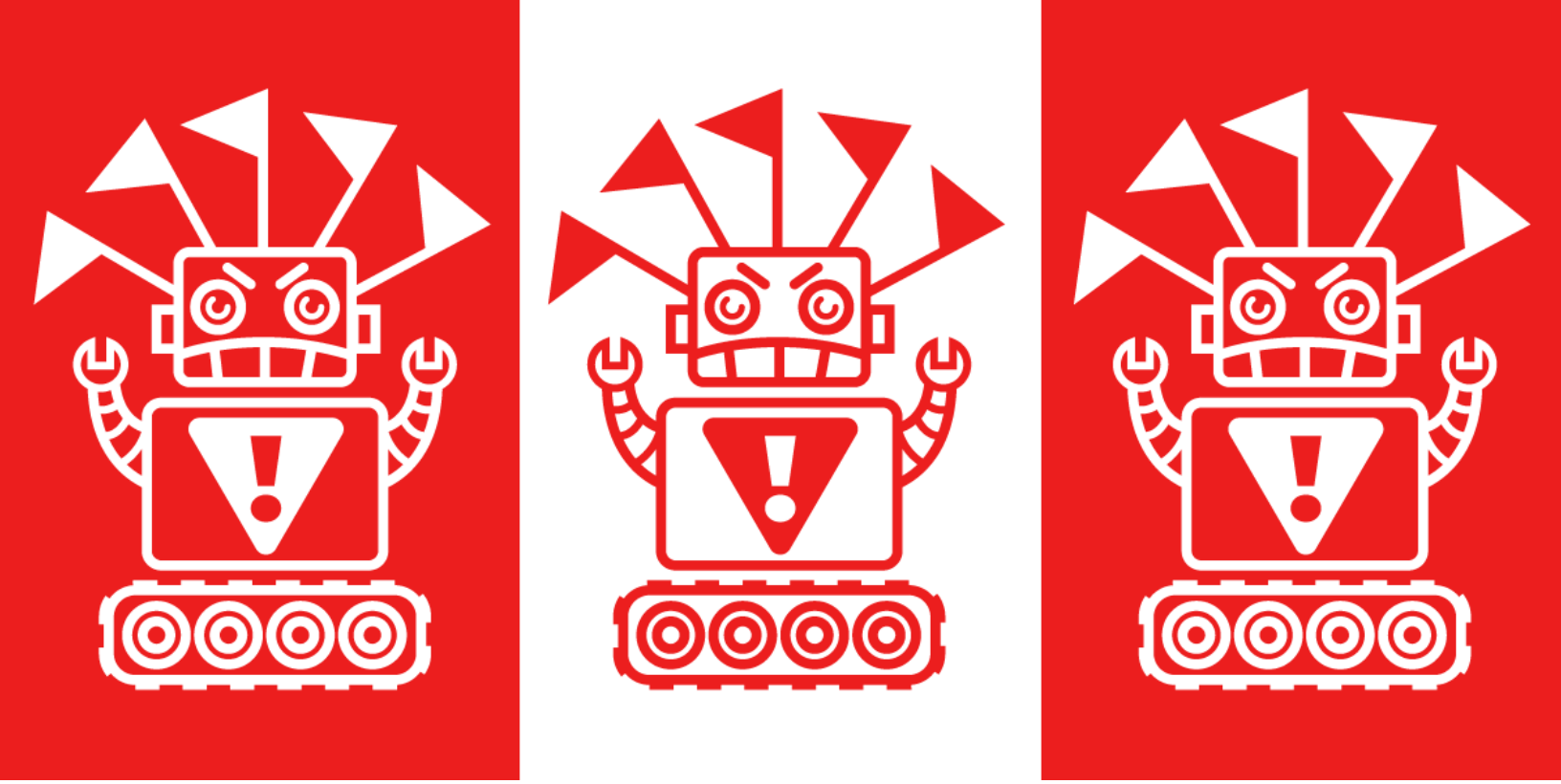


# GoGuardian:

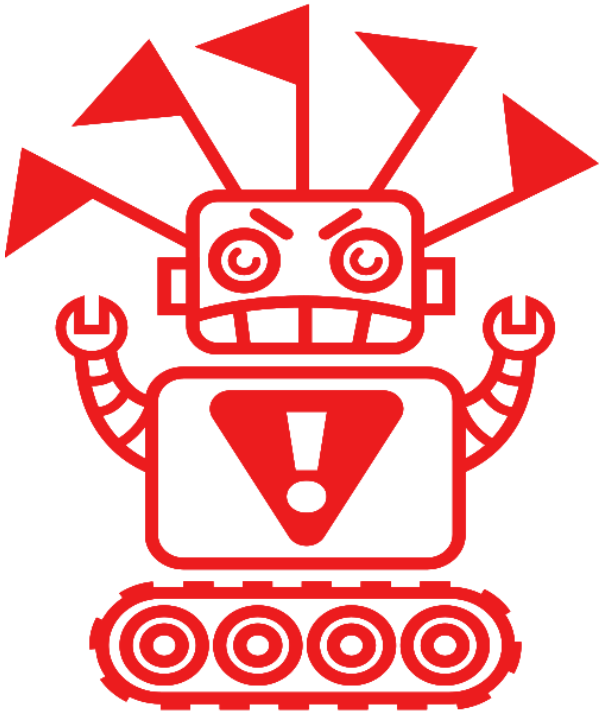
## A RED FLAG MACHINE BY DESIGN

ELECTRONIC FRONTIER FOUNDATION

**Authors:** Dave Maass, Daly Barnett, and Jason Kelley

A publication of the Electronic Frontier Foundation, 2023.


View this report online: www.redflagmachine.org



# GoGuardian:

## A RED FLAG MACHINE BY DESIGN

Dave Maass, Director of Investigations
Daly Barnett, Staff Technologist
Jason Kelley, Activism Director
Additional research by Beryl Lipton and Kieran Dazzo

**October 2023**

# Table of Contents

*Note: Due to the nature of the words that GoGuardian flags, this report and the Red Flag Machine quiz contain language that some may consider 'adult.'*

# Executive Summary

Nearly 88% of schools report that their district uses some form of student activity monitoring software. One of the most common is called GoGuardian. School districts that use GoGuardian software to monitor students' online browsing each receive thousands upon thousands of alerts every month based on keyword triggers. These triggers are, at least in theory, meant to alert school officials to content that students should not be viewing. Larger school districts, according to the company, may receive as many as 50,000 warnings a day that students may be accessing inappropriate content.

Are youth really spending that much time exploring the dark and dirty corners of the web?

No. Not at all.

According to datasets obtained by EFF, students mostly just want to do research for their homework, play games, and watch music videos. They want to indulge their curiosity, explore their identities, find part-time jobs, apply to colleges, and shop for prom dresses. Sure, they occasionally visit some PG-13 pages or listen to songs that would trigger Tipper Gore, but attempts to access pornography are actually relatively rare.

Yet, due to an over-inclusive dictionary and an undiscerning yet aggressive webpage-scanning formula, GoGuardian—and, we believe, other similar software—routinely misidentifies perfectly run-of-the-mill websites as potentially harmful or explicit.

For example, GoGuardian flagged a Texas student for visiting the online Texas Driver Handbook, which warns about cocaine's impact on concentration. The mention of "cocaine" was enough to generate an alert.

GoGuardian flagged another student for visiting the text of Genesis from Bible.com, because Adam and Eve are "naked" in Genesis 3:7-8.

The Texas House of Representatives' website was flagged because certain bills use the words "cannabis" and "sexual," such as legislation relating to regulation of medical mairjuana or the prosecution of sexual assault.

Poetry by the Brontë sisters, the text of George Bernard Shaw's play *Pygmalion*, even *Romeo and Juliet* all set off flags. And a Google search for "Moby Dick" was flagged for....well, you guessed it.

These are just a few of the tens of thousands of examples we reviewed. EFF obtained these examples from 10 schools in both red and blue states, including California, Florida, New Mexico, Rhode Island, and Texas. The documents reveal what websites are "whitelisted" and "blacklisted,"and which website visits triggered alerts, based on which keywords. We repeated the exercise a few months later after GoGuardian claimed to have adjusted its keyword dictionary. There was little discernible improvement.

The majority of the flagged websites are unobjectionable, because GoGuardian's process necessarily produces a flood of false positives. Yet the company insists these false positives are a feature, not a bug, even though educational tools, employment listings, college recruitment sites, and health articles are caught in the dragnet.

In its user guides, GoGuardian acknowledges that its software creates "unnecessary and often times innocuous noise" and flags keywords that don't even appear on a webpage, but are buried deep in the source code and metadata and are "not necessarily being searched for intentionally by a student."

GoGuardian functions as a red-flag machine by design. Rather than carefully determining whether students' online activity is truly harmful or explicit, GoGuardian pumps out flags like a factory line—so many that their marketing materials have to work overtime to make it sound like a good thing. Its over-inclusivity is seen as a plus, even as it forces administrators to comb through the data themselves to separate the genuine threats from the bogus leads. Its flags are overbroad, inaccurate, and potentially dangerous to students.

To illustrate the shocking absurdity of GoGuardian's flagging algorithm, we have built the Red Flag Machine quiz.

GoGuardian isn't the only student monitoring tool we're worried about, of course.

K-12 school districts around the country have begun adopting these new online monitoring technologies in large numbers, with the stated goal of preventing suicides, shootings, bullying, self-harm, and drug abuse among students, as well as blocking access to pornography, explicit music and videos, and video games. Many districts use these tools in an effort to comply with the Children's Internet Protection Act, or CIPA, which requires that schools which receive E-Rate funding must "block or filter Internet access to pictures that are: (a) obscene; (b) child pornography; or (c) harmful to minors," as well as requiring the monitoring of the online activities of minors.

But as reasonable as the intentions might be, the impact of this mass surveillance may ultimately harm the students they aim to protect. GoGuardian is one of the most common products, purchased by hundreds of schools nationwide, according to information obtained through the service GovSpend. Other common student monitoring tools include Bark, Securly, Gaggle, and Lightspeed Systems.

We've written a detailed description of GoGuardian here, but in short: GoGuardian allows schools to create lists of websites and domains that should be permanently

blocked on school-managed devices, as well as websites that should be allowed, even if they would normally trigger a block. It allows teachers and administrators to see in real time what students are looking at on their screens during class, and to close non-relevant windows. But it also includes a feature that scans every website visited by a student on a school-issued device, sometimes even when they are at home (and may not be the only user). And school-issued devices are not the only ones scanned: Students who are using personal devices over the school's internet connection, or those who connect their personal devices to their school-issued device, may also find their personal device scanned, as well.

"GoGuardian searches through the content and metadata for every website visited on a monitored device to help catch inappropriate content that may slip past traditional filtering," GoGuardian writes in its support guide. "View high-level lists of flagged sites for your entire organization or deep-dive into individual user activity to see exactly who was accessing the flagged content to help you better determine how to respond."

GoGuardian generates at least two different types of website flagging: "Flagged Activity" and "Smart Alerts." When available we analyzed both. We also received a small selection of alerts generated by a separate GoGuardian product called Beacon.

Here's how the company describes these features:

> Flagged Activity is generated when words on a site's page source match one or more entries on the Flagged Terms list. In instances where words are flagged on a webpage, the specific word(s) flagged may not have been visible to the student as not all content of a page's source code is always visible. Comments sections of webpages are another case where flagged activity may pick up words that are on a page which may be removed later due to moderation/administration.
>
> Smart Alerts are created when GoGuardian's proprietary algorithm determines that a page has explicit content based on the page's text. Smart Alerts was built to reduce some of the unnecessary and often times innocuous noise that Flagged Activity can create based on our data, intelligently analyzing page data with context instead of matching specific terms. If our algorithm is 50% or more confident that the page contains explicit content, a Smart Alert will be generated.

The admissions made here are astounding and worth repeating. GoGuardian openly acknowledges that its flagging feature raises alarms based on words that weren't even visible to the student, such as text that appears in the underlying HTML or Javascript of a website. It also admits that this feature generates "unnecessary and often times innocuous noise." The floor of 50% confidence used by the Smart Alerts algorithm seems extremely low.

However, it seems like the Smart Alert feature may not be widely used. Of more than 10 school districts we queried, only one was using Smart Alerts. The rest only provided Flagged Activity data.

Analyzing this data, it becomes clear that GoGuardian is an example of software that turns the blessing of youthful curiosity and ambition into a curse. But it's time-consuming to comb through tens of thousands of URLs and keywords, so to make our point quickly, we made the Red Flag Machine Quiz.

# The Red Flag Machine Quiz

Derived from real GoGuardian data from three school districts, visitors are presented with websites that were flagged and asked to guess what keywords triggered the alerts. There are over 50 different examples, and after each quiz question we explain why the keywords were triggered, if we could figure it out.

Some of GoGuardian's flags are admittedly very funny. An eBay listing for a Cheetos snack shaped like a mushroom was flagged for "penis." A Google image search for "buff elmo" was flagged for "sexy" and "motherfucker." But the dangers of this intense, widespread surveillance are anything but a joke.

Ultimately, this type of software creates records that incorrectly accuse students of accessing or seeking out materials related to sex, drugs, weapons, self-harm, extremism, or other potentially dangerous activities.

In some cases, the alerts have been shared with parents or law enforcement, which can happen when they are issued outside school hours. Almost one in two teachers say that students have been contacted by law enforcement as a result of student monitoring software, according to a 2022 report by the Center for Democracy & Technology (CDT). Baltimore City Public Schools, for example, told The Real News, "Clinical supervisors, school police, principals and designated school staff receive all GoGuardian alerts," and "(i)n cases where the student is not in school and contact cannot be made, wellness checks are conducted by school police and follow up with school-based clinicians."

GoGuardian software may also have an outsized impact on students seeking out materials related to protected classes. CDT's research has shown that 48% of Black students and 55% of Hispanic students, or someone they knew, got into trouble for something that was flagged by an activity monitoring tool, while only 41% of white students reported having similar experiences. Students with individualized education plans and/or 504 plans (those who have a learning or other challenges that necessitates specially designed instruction that is documented and reviewed at least annually) are more likely than their peers to report that their online activities are monitored (89 percent versus 78 percent, respectively). LGBTQ+ students are also more likely to be disciplined, and licensed special education teachers report higher incidents of law enforcement contact among their students.

We found these results unsurprising. GoGuardian disproportionately flags music and literature by Black creators because it makes no attempt to distinguish between hate speech and artistic expression. It also has flagged historical descriptions of slavery. It flags research related to Jewish experiences during World War II and the Holocaust for the use of the word "Nazis." And it flags explanations of sexual discrimination because, to GoGuardian, all discussions with the word "sexual" in them, even those meant to help young people protect themselves, are potentially harmful.

The software also exposes students searching for LGBTQ+ resources. In at least one 2022 example we saw, the software tracked a student who searched "am i gay test" on Google, then visited an online novelty quiz. As a result, GoGuardian generated an "incident" report for the student's consumption of "explicit" content. In fact, until early 2023, GoGuardian specifically flagged any website that included the word "lesbian." But that update didn't solve the problem; we saw examples of the same "am i gay" quiz flagged in 2023, after the change, for containing the words "sex" and '"sexual."

CDT's research also shows that LGBTQ students are more concerned than other students about this level of tracking, in part because 13% of all students in schools using this type of software have themselves, or know someone, who hasbeen outed by it. And much content isn't just flagged: One-third of teachers report that content associated with LGBTQ+ students or students of color is more likely to be filtered or blocked.

It's not only the "flags" that are cause for concern, of course. This type of monitoring software often records *all* websites a user visits, whether they are flagged or not—making it easy for administrators and teachers to sift through them. Students, and parents, likely don't realize that this level of private detail is available to administrators. In one 2022 instance, we could reasonably determine that a particular student was dealing with irritable bowel syndrome. In more recent results, we saw flagged results for gender dysphoria, various cancer treatments, and pharmacy information.

You can get a quick sense of how outrageous GoGuardian's flags are by taking the Red Flag Machine quiz, which includes dozens of examples of flagged sites. The report below is a more comprehensive view of the problem, its impacts, and how to fix the issue. It outlines our extensive findings, and also provides examples of these data sets.

# Methodology

The Red Flag Machine quiz and this report are primarily based on datasets obtained by EFF through public records requests filed with school districts under state-level freedom of information laws in 2022-2023. Our understanding of GoGuardian's functions is based on materials available through the company's public "Unified Help Center."

We started by using the proprietary online service GovSpend, an aggregator of purchase orders and other procurement data, to identify school districts that have entered

contracts with GoGuardian. We selected 27 school districts representing a variety of regions and sizes, with a particular focus on comparing school districts in states with new restrictions on LGBTQ+ and race-related content (e.g. Florida, Texas) against school districts in California, which has taken a polar-opposite stance on these issues.

We used the public-records tools at Muckrock News to file records requests with these 27 districts. We sought a variety of datasets, including: what domains or websites have been permanently blocked ("blacklisted") or allowed ("whitelisted") and all alerts ("Flagged Activity" and "Smart Alerts) that were generated in response to student browsing through GoGuardian's various filters. We also asked for alerts generated through another GoGuardian tool called "Beacon." The initial round of requests sought data from 2020-2022, to coincide with the COVID-19 pandemic that resulted in an increase in remote learning across the country.

Less than a third of the school districts provided data in response to our request. Several districts did not respond to our requests at all. A handful of districts rejected our request, claiming privacy or security-related exemptions under public records laws. Some school districts acknowledged they used GoGuardian for certain administrative tasks, but did not use the web monitoring feature and therefore did not have responsive records.

Eight school districts provided data in response to our first round of public records requests: Alvord Independent School District (Texas), Fruitvale School District (California), Imperial Unified School District (California), Lake Travis Independent School District (Texas), Las Cruces Public Schools (New Mexico), Lincoln Public Schools (Rhode Island), Southwest Independent School District (Texas), and Union County School District (Florida). Two additional agencies, Mount Prospect School District 57 (Illinois) and Dublin Unified School District (California), provided partial records.

In most cases, the records only went back five months, which seemed to be a limitation of the software. Nevertheless, the "Flagged Activity" datasets were huge, taking the form of spreadsheet files or print-to-PDF versions of the spreadsheets. The "Flagged Activity" data sets were often massive files. In fact, GoGuardian seemed unable to export more than 10,000 data points at a time, so some of the datasets we received were incomplete.

Each "Flagged Activity" dataset included fields for email and IP addresses, which were redacted or deleted to protect the students' privacy. The remaining fields included the URL, when the URL was accessed, what keywords triggered the software, and how many times those keywords appeared.

In one case, a school district provided not only the flagged data, but what appeared to be all browsing data, site-by-site. The agency failed to adequately redact the email addresses of the individual students, inadvertently revealing sensitive information traceable to individual students. EFF immediately contacted the district about the problem and had the staff re-redact the data. However, the incident illustrates how easy it is for school administrators using this software to make a mistake and violate the

privacy of all students in their care. It also shows how alarmingly easy it is to pry into the private lives of students and track their every online inquiry.

We reached out to school districts with a series of questions and an opportunity to provide comment and context, but only Southwest Independent School District in Texas provided any responses—brief, often only one word—to our questions.

In March, we contacted GoGuardian to confirm that our understanding of the software's features (and failings) was accurate. We noted two of our concerns: First, the software flagged non-English words like "pico" and "huevos," which may be slang, but also resulted in food items such as pico de gallo and huevos rancheros on restaurant menus being flagged; second, we asked the justification for including several words that relate to LGTBQ issues, like "lesbian" and "homo." A GoGuardian representative informed us that the keyword list had been updated recently:

> As part of GoGuardian's ongoing product improvement cycle, we regularly review and assess our features. Following our most recent review of Flagged Activity (Q4 of 2022), the keyword list includes only terms directly linked to the school's requirements under CIPA regarding explicit and harmful content. The current list does not include any LGBTQ terms or terms in languages other than English.

The GoGuardian representative provided us with dates that would be more representative of the adjusted wordlist.

Accordingly we filed new requests with the eight school districts. Five agencies that previously provided the records did not fulfill the request this second time around. Lincoln Public Schools provided the data as a PDF, but the URLs were not legible and the district did not respond to requests for the data in an alternative format. The Union County School District and Southwest ISD both provided spreadsheets, but had decided to delete the URL categories, which means we could view the keywords but not the webpage where the software reportedly found the words. Interestingly, Las Cruces Public Schools and Fruitvale School District said they each had ceased using GoGuardian since our previous request

We received full datasets from Alvord ISD, Lake Travis ISD, and Imperial USD, which serve as the basis for the Red Flag Machine quiz. However, this report and analysis also draws from the previous datasets we received, since the overwhelming majority of keywords did not change between the two periods.

EFF is not making the datasets publicly available at this time due to concern that the data may unintentionally reveal information that could be connected to an individual student. However, researchers and journalists seeking to verify our findings may contact us at dm@eff.org and we will make a case-by-case decision about sharing the data.

# Red Flags Everywhere

In a 2021 letter responding to concerns submitted by Sen. Elizabeth Warren and Sen. Edward Markey, GoGuardian provided some staggering claims about the aggregate number of alerts generated by its Admin product. At the time, GoGuardian said the software was used at 6,700 schools and school districts, and accounted for more than 9.5 million monitored accounts—which we take to mean it watched over 9.5 million students.

"Over the course of 2020, Admin generated 44 million alerts or approximately 4.6 alerts per student over 2020," GoGuardian CEO Advait Shinde told the senators. "Of these alerts, approximately 90% were for explicit or inappropriate content and approximately 10% were for self-harm."

Shinde's comment here is misleading. While GoGuardian may have generated 40 million alerts that the software *tagged* as explicit or inappropriate, it is not true—not by a long shot—that most or all of the content was actually explicit or inappropriate. Shinde did not tell Warren and Markey the part about the software generating "unnecessary and often times innocuous noise" or that "a search for cats may cause what may appear to be inappropriate flagged activity"—both of which are direct quotes from the GoGuardian support guide.

When it comes to school safety, GoGuardian is a blunt instrument, not a scalpel: It errs, so to speak, on the side of error. By generating thousands of false positives, GoGuardian leaves school administrators to sift through enormous haystacks to identify a needle of truly harmful content. According to GoGuardian, "some larger school districts can generate upwards of 50,000 alerts per day."

While the justification for the flagging may be to identify inappropriate content, at any given school, a large number of results are for something far less explicit: either direct links to video games or attempts by students to circumvent school filters to access video games. For example, of the 2,313 websites flagged by GoGuardian at Alvord ISD between February 15 and March 22, 2023, 14% were flagged for the keyword "unblocked." At Imperial USD, more than a quarter of the 15,000 websites flagged by GoGuardian between March 10 and March 17, 2023 were for the keyword "unblocked." In both cases, virtually all of the visits were related to students trying to access games.

Other common types of non-explicit sites that GoGuardian regularly flagged included college application sites and college websites; counseling and therapy sites; sites with information about drug abuse; sites with information about LGBTQ issues; sexual health sites; sites with information about gun violence; sites about historical topics; sites about political parties and figures; medical and health sites; news sites; and general educational sites.

As a case study, we can look at the Imperial USD in Southern California, which has approximately 4,400 students. In just six days (March 17-22, 2023), GoGuardian flagged 9,387 website visits. Of those:

- More than 2,302 are Google searches, including:
    - More than 1,000 for unblocked video games (74 alone for a game called "Slope")
    - 72 for examples of political cartoons about issues such as limited government, bureaucracy and gun violence
    - 44 for examples of protest signs, slogans, and chants
    - 41 searches for jobs
- 2,400 visits to YouTube, including:
    - 548 YouTube Shorts, usually tame TikTok-style short viral videos
    - 107 visits to the YouTube.com front page
- 2,000 visits to Sites.Google.com domains related to unblocked games (Minecraft, Retro Bowl, and Five Nights at Freddy's being among the most popular)
- 427 visits to Spotify
- 260 visits to the website Coolmathgames.com
- 157 visits to Wikipedia pages
- 83 visits to an episode of RadioLab segment about beauty
- 68 visits to the Imperial School district's Instructure online-learning platform
- 46 visits to Rhymezone.com, a site for finding words that rhyme with other words
- 44 visits to the United States Holocaust Memorial Museum's website

There were a little more than a hundred searches overtly related to drugs, although some were about the risks and dangers of marijuana or the film "Cocaine Bear." We found no direct references to suicide (with the exception of rapper Ken Carson's song *Suicidal*).

Browsing related to pornography was minimal during that week-long period at Imperial USD. A few searches ranged from clear attempts to view hardcore pornography on reddit to just curiosity over what it looks like for two sabretooth tigers to mate. Often when a student searched for explicit material, it was related to animated characters such as Minnie Mouse and the Gem characters from *Steven Universe*. However, the data does show that on two particular days, a student (or students) used Bing *after school hours* to view dozens of sexually explicit illustrations, mostly from anime or manga. However, on the whole, these searches represented a fraction of a fraction of student browsing.

# Causes of Incorrect Flags

## An Over-Inclusive Dictionary

When GoGuardian scans a webpage, it compares the text against a long list of keywords. While some of these keywords are obviously explicit (e.g. "MILF," "fisting," etc), many are so broad that they should reasonably be expected to generate (mostly) false positives.

For example, in the Lake Travis ISD dataset from February and March 2023, more than 900 website visits were flagged for the term "colon," which we assume GoGuardian considers explicit because of its proximity to other parts of the body often involved in sexual activity. However, this resulted in students being flagged for visiting pages about how to use a colon or semicolon in sentences and mathematical formulas, biographical pages about Christopher Columbus (aka Cristóbal Colón), and educational pages about human anatomy.

The word "cox" was also a keyword, probably because it's an alternative spelling for "cocks." It resulted in the common surname "Cox" being flagged. Similarly the term "Wang" was flagged, even though most search results were related to someone's surname and not the slang for "penis."

"Nazi" was likely included in the dictionary to flag websites containing extremist content or hate speech, but this also resulted in many students being flagged for researching World War II.

Anything including the term "sex," including biology-related content, was flagged, as was clothing that's described by the seller as "sexy." The word "hardcore" was flagged regularly despite it usually referring to video games (Minecraft has a "hardcore" mode), or in some cases, in reference to sports (as well as John Belushi, and life at Auschwitz). And some very common words are included that generally aren't used in explicit ways, like "hot" and "hottest," as was the case in the Alvord ISD dataset.

## Synonomy

A common, accurate critique of content moderation—which is one way to consider what GoGuardian is doing—is that it is impossible to do well at scale. A major reason is the problem of synonymy—many words are multi-dimensional. As we mentioned above, flagging the keyword "Nazi" is likely going to produce an enormous number of false positives: It *might* signal that content is related to white nationalism, but it can also be used as an accurate descriptor (not to mention its frequent, if less accurate, derogatory use). If Facebook blocked all content that contained the word "porn," it would also be blocking content that was anti-porn, because to describe the thing you often must mention it.

This is a regular issue for content moderation on social media sites, and it comes up in GoGuardian flags often. Some other examples we saw included multiple pages about the benefits of abstinence flagged for containing the words "vagina" and "penis," therapy sites flagged for discussing how to deal with suicidal ideation; educational sites about drug abuse flagged for including the names of drugs; and news stories about the March for Our Lives, a series of student-led gun control demonstrations, were flagged for mentioning guns.

These types of flags were sometimes so inaccurate as to be darkly ironic. GoGuardian flagged a PEN America report called "Banned in the USA: The Growing Movement to Censor Books in Schools" not because it contains content that is dangerous, but because it mentions terms that censors use to describe some books that have been banned in the past, including 'porn' and 'sexual.' The "Censorship in China" Wikipedia page, part of a Wikipedia series on censorship in various countries, was flagged, in part, because the words "sexual," "naughty," and "porn" were all descriptions of content that has been censored in China. And the word "Nazi" is included as a link to censorship under Nazi regimes.

# Comments, Recommended Links and Source Code

GoGuardian openly acknowledged that it flags content that may not be viewable by the student or that the student did not intend to view in the first place.

## Comments

GoGuardian flagged many websites not for the content of the page, but for content that users had posted to the "comments" section, particularly on articles at the center of online controversies. For example, GoGuardian flagged the below two posts not due to the article, but because of the heated "flat earth conspiracy" discussions in the comments.

- An article titled "What Would Happen if the Earth Were Actually Flat?" on the Columbia University Climate School's website was flagged for terms like "anal," "screwing," and "suicide," which appeared only in the comments.
- A post on Scienceblogs.com answering the question "Who Discovered The Earth is Round?" was flagged for "ass," "fucked" and "suicide," which only appeared in the comments.

But in some cases, the websites were flagged because online spammers had targeted the comment section. For example, GoGuardian flagged a Lake Butler community website's photo gallery of honor roll students from Kindergarten through 4th grade. Why? The website administrators had not turned off the "pingback" feature in the comments, resulting in 521 spam links, mostly related to pharmaceuticals (including ones for erectile dysfunction).

### Source Code

Because GoGuardian scans the underlying HTML source code of web pages for keywords, it will often flag terms that are not viewable to the student. For example:

- GoGuardian flagged the product page for Purina pig food on TractorSupply.com because the terms "butt" and "fecal" appeared in the source code.
- A Walgreens page showing the hour and location for the Bee Cave, Texas branch was flagged because the term "hardcore" appeared in the url of an image in the page source.
- An Epilepsy Foundation article about the benefits of service animals was flagged because the words "cannabis" and "rectal" appeared in the page source as headlines to other, legitimate articles on the website.
- A National Archives page displaying scans of the U.S. Constitution was flagged for "bloody" and "Nazi" because the book titles "The Impact of Bloody Sunday in Selma" and "Life in Concentration Camps: The Horrors of Nazi Germany" are in the Javascript in the viewable page source.

It's worth noting that, especially in the latter case, not only were the keywords "bloody" and "nazi" not visible without viewing the source code of the page, but even then, they were links to academic content.

### Recommended Links

The "World Wide Web" is not just an analogy, but a description—the "web" is built on links, and often that means content, previews, and menus on a site may contain descriptions or names of other pages that are not located on the flagged page itself. The Constitution example above, where names of other documents offered on the site were in the source code, is one example, but there were many others. GoGuardian often flagged news sites for this. A student reading about a kitten rescued from a tree may have been flagged because a link to another story about cannabis regulations was also on the site; a Labroots.com article titled, "Here's What Would Happen if You Fell Into Saturn's Atmosphere" was flagged because "Cannabis Sciences" is among the "Trending Categories" in the sidebar.

# "Smarter" GoGuardian? Smart Alerts and Beacon

## Smart Alerts

Acknowledging that the standard flagging mechanism creates an enormous amount of alerts, GoGuardian offers a secondary feature called "Smart Alerts" that only generates

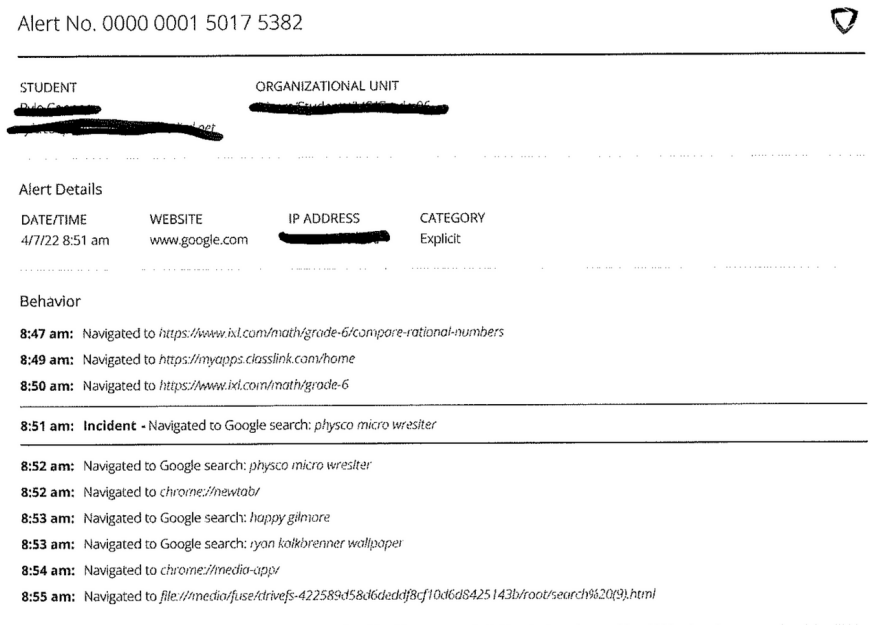
alerts if an algorithm decides with 50% or greater confidence that the browsed material is explicit.

Smart Alerts are described by GoGuardian as a "feature that uses advanced machine learning algorithms to continuously monitor student browsing activity for inappropriate content." The software scans in real-time to generate alerts.

However, a sample of Smart Alert documents provided by the Alvord ISD in Texas–show that even this feature is frequently wrong, despite the claimed 50% or greater confidence threshold. (Only one other school district—Union County School District in Florida—provided Smart Alert data, however it was not presented in a format that allowed for granular analysis.)

The Alvord ISD documents, provided as PDFs, show that not only does GoGuardian flag individual pages searched by a student, it also tracks the online path that a student follows to arrive at a page with suspected explicit content. Each alert contains the student's name and IP address (both redacted) and a "behavior" section that includes the browsing history leading up to and following an "incident." Each step is time-stamped.

For example, one alert shows a student working one morning in April on some math tutorials. They appeared to take a break, and over the next four minutes they used Google to look up a professional micro wrestler named Syko, the Adam Sandler movie *Happy Gilmore*, and desktop wallpaper featuring college basketball star Ryan Kalkbrenner. It was the wrestler's name, misspelled as "physco" by the student, that triggered an "explicit" alert.



Alert No. 0000 0001 5017 5382

STUDENT [redacted]

ORGANIZATIONAL UNIT [redacted]

Alert Details

| DATE/TIME | WEBSITE | IP ADDRESS | CATEGORY |
|---|---|---|---|
| 4/7/22 8:51 am | www.google.com | [redacted] | Explicit |

Behavior

**8:47 am:** Navigated to *https://www.ixl.com/math/grade-6/compare-rational-numbers*
**8:49 am:** Navigated to *https://myapps.classlink.com/home*
**8:50 am:** Navigated to *https://www.ixl.com/math/grade-6*
**8:51 am:** **Incident -** Navigated to Google search: *physco micro wresiter*
**8:52 am:** Navigated to Google search: *physco micro wresiter*
**8:52 am:** Navigated to *chrome://newtab/*
**8:53 am:** Navigated to Google search: *happy gilmore*
**8:53 am:** Navigated to Google search: *ryan kalkbrenner wallpaper*
**8:54 am:** Navigated to *chrome://media-app/*
**8:55 am:** Navigated to *file://media/fuse/drivefs-422589d58d6dedef8cf10d6d8425143b/root/search%20(9).html*

Innocent online video games regularly generated Smart Alerts, including games like a bow-and-arrow shooting gallery, a pirate battle royale game, a game where you drag colored strings across the browser window, and various educational and math-based games and activities.

Other innocuous behavior that generated an "explicit" Smart Alert:

- Searching for advice on stopping thumbsucking
- Trying to find out the name of actor Jamie Foxx's wife
- Googling the word "hot"
- Researching why Colombian people hold candles during traditional cumbia dancing
- Looking up information about the Dragon Ball Z character Goku
- Using a full-screen online calculator
- Reading "Tuesday Siesta," a short story by Gabriel García Márquez
- Misspelling "intramural sports" while trying to look up its definition
- Hunting down a Miss Manners quote

It's hard to imagine how an algorithm would identify a calculator as explicit with any percentage of confidence, let alone greater than 50%. In fact, the only actual explicit results out of the dozens reported seemed to be for a student who viewed the lyrics to Cardi B's WAP and a student who searched for "girl boobs" on Google.

On the other hand, the Smart Alerts did accurately identify a small number of occasions where a student was browsing content related to self-harm, such as one student looking up "how to sueside [sic]" and another researching whether people who commit suicide go to heaven.

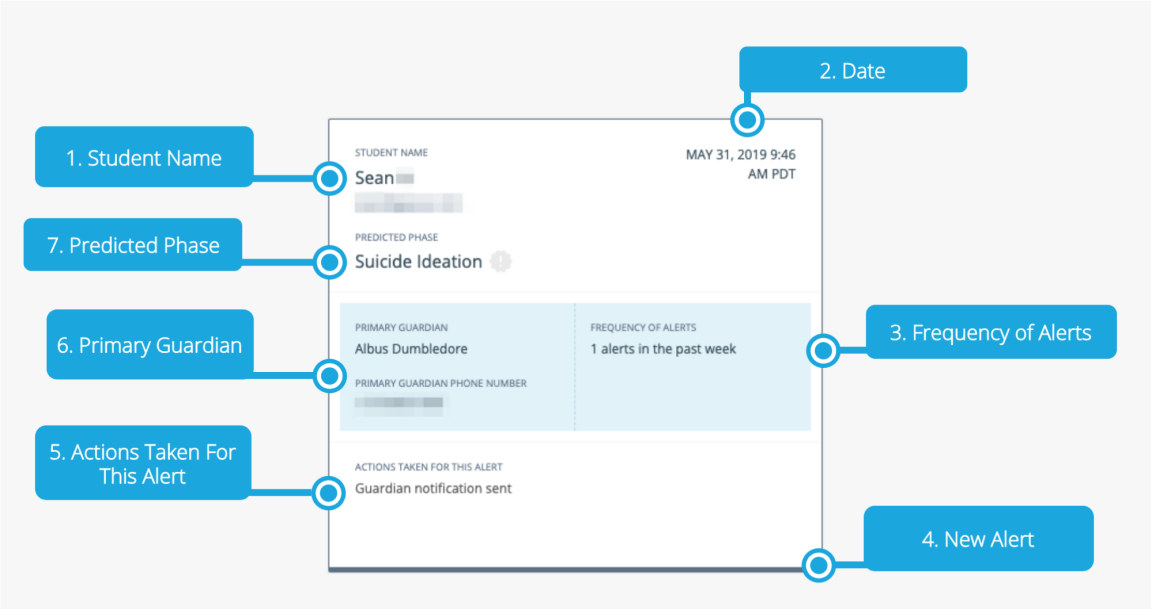
However, sometime in 2022, the "self harm" category was eliminated for new customers, leaving only "explicit" alerts.

## Beacon

GoGuardian now offers another product suite called Beacon, designed specifically for monitoring suicide and self-harm risks. Here's how it's described:

> Beacon is a machine-learning solution that notifies pre-determined school staff of instances where students may be at risk of suicide, self-harm, or potential harm to others. Beacon was designed to help school staff proactively identify at-risk students and quickly facilitate a response.
>
> Beacon is a multi-class classifier machine learning model built to identify students' online behaviors that could be indicative of suicide or self-harm by analyzing their browsing across all student content. This includes search

engines, social media, email, web apps, and more. Beacon shows a holistic view of the student's online behavior.




Only two schools provided Beacon Data: Union County, which recorded a couple dozen alerts, and Alvord ISD, which produced only one alert.

All of Union County's alerts listed the student as in the "Active Planning" phase, and indeed the majority of the alerts involved Google searches for obviously suicidal phrases like, "how to kill yourself" and "how much focalin is needed to overdose." (In addition to "Active Planning," other Beacon phases are Suicide Ideation, Self Harm, Help & Support, and Suicide Research. Given that GoGuardian categorized all of the Beacon results we saw as being in the "Active Planning" phase, we do not have any information on the accuracy of Beacon in determining these phases. But one imagines that "how much focalin is needed to overdose" could just also be considered "Suicide Research.")

Alvord ISD only provided one example of a Beacon alert. The incident involved a student searching the phrase "drive to die" in Google. Most of the results link to various video games in which the player drives a vehicle through a zombie apocalypse, while none relate to suicide.

One conclusion that can be drawn from this is that the self-harm filter was more effective at identifying students potentially in crisis because the terminology is very specific and clear, such as "I wanna die," which was the subject of several Beacon alerts in Union. Googling "suicide help hotline" is an obvious indicator that a student is seeking help for themselves or someone they know, as was the case with an Alvord ISD Smart Alert.

But even when GoGuardian is accurate in determining the content of the alert, and potentially even the phase, it is important to note that unexpected interventions resulting from surveillance of behavior are not always beneficial. If a school intervenes

unexpectedly when a student searches for a suicide hotline, for example, that student may not feel comfortable searching for such a resource the next time. Once students *do* know that searching for information online about their suicidal thoughts might get them a call from the guidance counselor, they may feel *worse* about the situation, not better. And depending on who learns of the student's Beacon alerts and reaches out to them, interventions could even cause more harm than good, particularly if a student has no connection with the person. Unfortunately, not all school officials—or parents—respond appropriately to learning a student has researched suicide.

In 2018, Facebook received significant criticism for introducing an algorithm to recognize posts that might signal suicide risk, then reviewing those posts and contacting local authorities if a user seems at risk. In addition to privacy concerns and uncertain legal requirements, a major concern was the lack of data on the result of these interventions. Likewise, GoGuardian offers no public data on the impacts of Beacon interventions. (Some student monitoring tools do offer statistics, though they are impossible to confirm. Securly, for example, lists the number of lives it has saved on its homepage: 2,076, as of this writing, with no explanation or data to support that number.) Facebook's suicide detection algorithm was banned in the EU, and it's unclear if their outreach program is still in effect.

All these concerns are assuming Beacon gets it right. In our datasets, Beacon was still prone to error, reporting several students for actively planning suicide who clearly weren't, such as one student who was using a crossword puzzle solver site to try to find a synonym for "Western Necktie" (the site responded with "bola" and "noose"). One student, presumably in anticipation of GoGuardian flagging their browsing, actually typed the search term, "how to tie a simple noose knot (NOT SUICIDAL) chill out" into Google.

# Impact on LGBTQ Content and Content Regarding Historically Oppressed People

## LGBTQ+ Terms Removed, Threat Remains

After our initial records requests in 2022, GoGuardian informed us that it had removed two sets of keywords: foreign language slang and LGBTQ-related terms. Gaggle also dropped LGBTQ terms from their keyword lists. But the software still flagged significant amounts of LGBTQ content in the 2023 results, for keywords like "sex," "breasts," "penis," and "vagina."

Misclassifying content related to LGBTQ+ topics as adult or inappropriate is a problem as old as AI and content moderation. Software designed to crawl over a website to determine if it is fit for a red flag will rarely have the nuance to determine when anything that mentions sexuality or gender identity is educational, artistic, pornographic, or otherwise. Without the context of in-real-life bigotry, anti-trans legislation, and recent shifts towards criminalizing abortion, these types of software

mistakes might seem like small snafus. But in the current political reality, they can create a real danger to students who are already under threat. The dangers posed by these student monitoring software are worsened by the recent increase in laws outlawing speech related to healthcare and LGBTQ+ issues.

In its response letter to the senators' investigation of the harms that GoGuardian and student monitoring apps like it pose to students from marginalized backgrounds, GoGuardian stated several times that "to protect the privacy of students, Admin and Beacon do not collect student-level demographic data. This means that GoGuardian cannot link a student's online activity to individual student-level demographics."

This not only misses the point, it contradicts what we saw in the records we reviewed. Tracking individual student activity directly to the student obviously reveals this information—to administrators, teachers, parents, and sometimes, law enforcement—if not to GoGuardian itself. In several of the records we reviewed through public records requests, we saw fields in the reports that showed student names, emails, and other identifying points.

The sheer amount of Wikipedia articles alone related to sexuality or gender identity or anything at all vaguely related to LGBTQ+ ideas in all of the reports we reviewed is testament to how naive and clumsy GoGuardian truly is. It is not only accusing students of inappropriate behavior for wanting to learn more about LGBTQ+ issues, but it puts them into the line of sight for school administrators and anyone with access to the admin dashboard of that GoGuardian implementation, effectively outing them to that staff.

For example, between February 15 and March 17, 2023, GoGuardian software in the Lake Travis ISD flagged more than 75 websites with the terms "transgender," "LGBT," "gay," "homosexual," "non-binary" or "queer" in the URL. The following table includes examples, along with the keywords that triggered the alert.

| School District | What Was Browsed | Flagged Terms | Notes |
|---|---|---|---|
| Lake Travis ISD | A page on MakingQueerHistory.com about Oscar Wilde | pedophile, sex, sexual | |
| Lake Travis ISD | An academic paper titled, "Family Rejection as a Predictor of Suicide Attempts and Substance Misuse Among Transgender and Gender Nonconforming Adults." | sex, sexual, cannabis | "Cannabis" only appeared in the page source, as part of the title of a related paper. |
| Lake Travis ISD | An article about "Title IX's | sex, sexual | |

| | | | |
|---|---|---|---|
| | Implications for the LGBTQ+ Community." | | |
| Lake Travis ISD | A San Diego Gay & Lesbian News article about the low regret rate for gender-affirming surgery. | sex, sexual, clitoris, penis | |
| Lake Travis ISD | A magazine article about expressing gender identity through fashion | nude, wang, sexy, sex | The page references the fashion designer Alexander Wang. |
| Lake Travis ISD | The Wikipedia page for the Transgender Rights movement | Bitch, sexy, cox, rape, sex, sexual | |
| Lake Travis ISD | A Planned Parenthood resource about transitioning | Butt, sex, penis, sexual, scrotum | |
| Lake Travis ISD | A Wikipedia page on the portrayal of transgender people in film. | Cox, vagina, nazi, sex, sexual, rape, sexy | |
| Lake Travis ISD | A Medical News Today article about the difference between the terms "transgender" and "transsexual." | sex, sexual, viagra | The page has an advertisement for Viagra. |
| Lake Travis ISD | A New York Times article about transgender people and employment discrimination | cocaine, rectal, sex, sexual | “Cocaine” and “rectal” do not appear to be on the page. |
| Lake Travis ISD | An article in The Guardian about transgender history | sex, penis, fanny | “Fanny” refers to persecuted individual Fanny Park |
| Lake Travis ISD | A YouTube search for TED talks about transgender issues | cocaine, sex | Cocaine does not currently appear on the page or in the page source. |
| Lake Travis ISD | An article about queer coding in Bram Stoker's Dracula | penis, sexual, sex | |
| Lake Travis ISD | The Wikipedia entry for the Gay Men's Chorus of Washington, DC | nazi, rape, sex, sexual | |
| Lake Travis ISD | An article about the "heartwarming" gay storyline in the show *Little Voices* | orgasms, sex, sexual, sexy, rimming | Neither "orgasms" nor "rimming" currently appear in the page or page source, but may have appeared in a previously recommended |

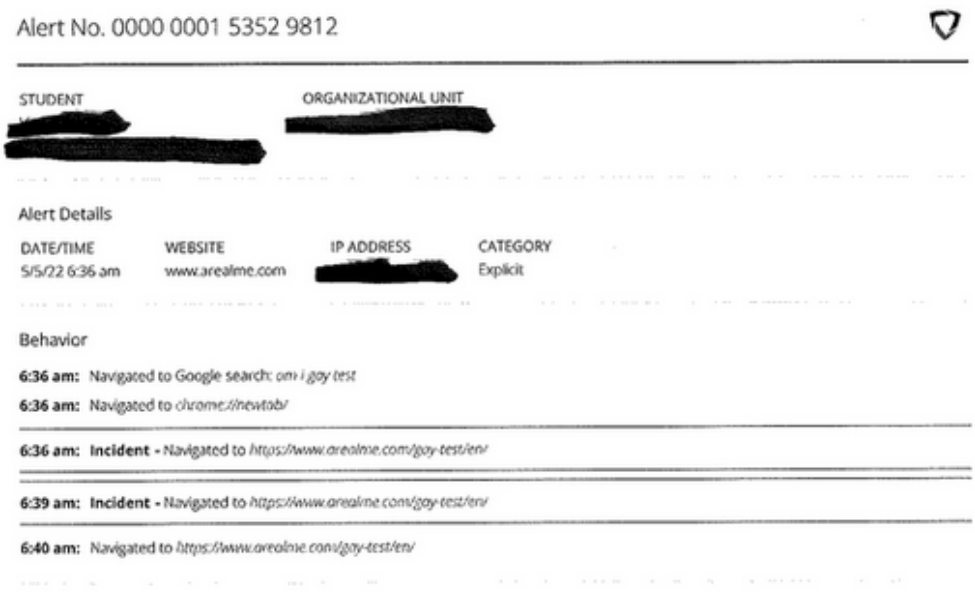
| | | | |
|---|---|---|---|
| | | | article. |
| Lake Travis ISD | An article about diagnosing gender dysphoria | penis, vagina, sex, sexual | |

Similarly, data from Imperial USD showed that simply searching for images using the term “trans people” on Google resulted in flags for “sex” and “sexual."

While pornography was not widely featured in the data, when it did appear, it was common for the pages to include non-heterosexual pornography, which could risk revealing a student’s sexual orientation.

One particular type of flagged website was common across multiple school districts: “Am I Gay?” quizzes. In particular, GoGuardian flagged a WikiHow quiz reviewed by a mental health clinician with a background in gender and sexual identity issues, and an arealme.com quiz, which was reviewed by an higher education professional, though the page includes a disclaimer that the quiz is for “for entertainment purposes only.” GoGuardian flagged the sites for using the terms “sex” and “sexual,” although in both cases, the term was employed in the context of sexual orientation. In addition, the “A Real Me” quiz was flagged because the word “testi” comes up in the page’s source code—in a Finnish translation of the page’s caption for “test and pass,” which in Finnish is, “testi ja selvit."

Flagging visits to these pages represents an extreme invasion of students as they navigate a complicated world of attraction, identity, labels, and stigma. In one example from Alvord ISD, GoGuardian’s Smart Alert algorithm elevated a visit to such a site to the attention of administrators.



Alert No. 0000 0001 5352 9812

STUDENT [redacted]

ORGANIZATIONAL UNIT [redacted]

Alert Details

| DATE/TIME | WEBSITE | IP ADDRESS | CATEGORY |
|---|---|---|---|
| 5/5/22 6:36 am | www.arealme.com | [redacted] | Explicit |

Behavior

**6:36 am:** Navigated to Google search: *am i gay test*

**6:36 am:** Navigated to *chrome://newtab/*

**6:36 am:** **Incident -** Navigated to *https://www.arealme.com/gay-test/en/*

**6:39 am:** **Incident -** Navigated to *https://www.arealme.com/gay-test/en/*

**6:40 am:** Navigated to *https://www.arealme.com/gay-test/en/*

The record shows that one morning in May 2022, before school opened, a student searched for "am I gay test" on Google, then opened "A Real Me" in a new tab, triggering an "explicit" content alert. Again, even though GoGuardian promises that "Smart Alerts" will occur only when the algorithm identifies explicit content with greater than 50% confidence, there is nothing particularly explicit about the quiz.

This incident is alarming given the rise of parental notification policies that require school administrators to alert parents if their students change their gender or pronouns, or in some cases, sexual orientation. For example, an Alabama law forbids school officials from withholding from parents "information related to a minor's perception that his or her gender or sex is inconsistent with his or her sex." It is unfortunately possible that an administrator could interpret a GoGuardian alert related to gender dysphoria to be such "information." Before it was halted due to a legal challenge, a school board in New Jersey passed a policy that would have required a school employee to inform parents "whenever the staff member is made aware of any facts or circumstances that may have a material impact on a student's physical and/or mental health and/or social/emotional well-being," with sexual orientation and gender identity and expression explicitly listed.

Other parental notification laws may require schools to alert parents when lessons or school materials discuss gender identity or sexual orientation. For example, Arkansas' law requires schools to alert parents if students will be accessing materials (broadly defined) that touch on those issues and allow parents to inspect those materials. Should a student choose independently to research LGBTQ issues online in the context of any

class assignment (and not just sex ed), a school could interpret the rules as requiring parental notification.

## Content Regarding Historically Oppressed People

In reviewing the data across all school districts that provided records, it became clear that GoGuardian flags have a distinct racial and ethnic bias, disproportionately flagging content related to subject matter relevant to historically oppressed peoples.

This bias manifested in a number of ways. As noted earlier, pages that discuss the history of oppression often must use terms on GoGuardian's keyword list to discuss atrocities and racist organizations, such as "rape," "bondage," and "Nazi." GoGuardian regularly flagged content related to the Holocaust and antisemitism, and content related to Latinx people, such as Wikipedia pages related to Mexican-American history and anti-Mexican sentiment as well as a Library of Congress Latinx resource. However, content related to Black people or Black culture were by far overrepresented in the flagged data. This table reflects some of the more egregious examples:

| School District | What Was Browsed | Flagged Terms |
|---|---|---|
| Alvord ISD | A Google Image search for "Black Power" | fists |
| Alvord ISD | A Khan Academy page about the Emancipation Proclamation | blood, sex, bondage |
| Lake Travis ISD | An article about Maya Angelou in The Guardian | rape, rapist, sex |
| Lake Travis ISD | The text of the Civil Rights Act of 1964 | sex, sexual |
| Imperial USD | The Wikipedia page for Martin Luther King Jr. | bobo, bloody, dick, sex |
| Imperial USD | The text of Zora Neale Hurston's "Their Eyes Were Watching God." | gun, shoot, n—-- |
| Lake Travis ISD | A Wikipedia page about "Black codes" used to police Black people in the U.S. | sexual, n—--, nazi, bondage |
| Imperial USD | A CNN video about a Black pastor being arrested by Alabama police while watering flowers | gun |
| Imperial USD | An article about the global context of the civil rights movement | Bloody, bondage, nazi, sex, sexual, n—-- |
| Imperial USD | A Google search for "a picture that represents malcom x" [sic] | Gun |

Perhaps most noteworthy was the sheer amount of music GoGuardian flagged: Black artists accounted for a huge number of music videos, songs, and lyrics that the software flagged. This is a reflection of inherent bias in GoGuardian's focus on explicit terms and how that conflicts with some genres of hip-hop musical tradition that involve reappropriating slurs or using explicit language to express the emotional realities and individual experiences associated with oppression, poverty, overcriminalization, sexualization, and marginalization. For example, of the 43 individual songs on Spotify flagged by GoGuardian in Imperial USD during a 30 day period, 85% were by Black artists, such as Ice Cube, Jay-Z, and Ice Spice, usually for including the N-word or B-word. In essence, GoGuardian's dictionary hews toward expansive and somewhat old-fashioned concepts of vulgarity and doesn't recognize what is now mainstream artistic expression.

In states like Florida and Texas, legislators are passing laws making it harder for students to gain access to information about many cultural histories and perspectives, often using the term "Critical Race Theory" or CRT as a bogeyman for a much broader censorship agenda. Black history so far, has proven among the earliest casualties of state censorship and attacks by the self-identified "parental rights" movement. As books are challenged and pulled off shelves, students can and will turn to the internet. However, tools like GoGuardian set up a framework for further monitoring and control of students seeking out diverse viewpoints and experiences through art, literature, and history. It is conceivable that challenges of books in school libraries could evolve into challenges of websites on the open internet—or worse, blockage of any content with particular keywords, such as "racial justice" or "gender dysphoria." If the technology exists, there is always the potential for abuse.

## Eliminating Foreign Language Slang

In our 2022 results, GoGuardian's dictionary pulled slang from foreign languages, including obscure, regionally specific slang from across Latin America. This resulted in a large number of false positives.

For example, "Huevos," the Spanish word for eggs, is sometimes slang for testicles. This resulted in GoGuardian flagging webpages that referenced the popular breakfast menu item, "huevos rancheros." Because "palo," the Spanish word for stick, was also flagged, students were flagged when they tried to get information about Palo Alto College.

After GoGuardian updated the keyword dictionary, we saw little evidence of these types of flags. Some other slang that is arguably foreign is still included, such as wanker, kenke, knob, and bellend, but overall, this was an improvement.

# How to Fix the Broken Machine

## GoGuardian Deserves Much of the Blame—and Can Do A Lot to Improve

As it stands, GoGuardian's flagging functionality is irreparably broken. It is clear from our research that scanning for keywords and flagging websites based on those results does not serve any real purpose. There is no accurate way for software to know whether the content of a webpage, a document, or a video is harmful or not, based only on a few keywords.

We note above that GoGuardian errs on the side of error, but GoGuardian *could* err in the other direction. For example, if a page is flagged for containing words like sex or sexual, the algorithm could then ask: Does it also contain keywords like "lesbian" or or other LGBTQ+ keywords? Does it include biology-related terms like "genetics?" If so, it should not be flagged. Perhaps the company could create a counter-algorithm that evaluates a flagged page for educational value, so that when something is flagged on a keyword, another algorithm is searching for hallmarks that it's an academic resource. This mistake is particularly common when flagged words have multiple meanings or uses—like "sex" or "nazi," for example.

According to GoGuardian, its keywords "include only terms directly linked to the school's requirements under CIPA regarding explicit and harmful content." Keywords like "unblocked," "tumblr," "proxies," "poppy playtime," and "colon" certainly don't fall into this category. Whether "masochist," "ejaculate," or "s.o.b" do may be up for debate, but a counter-algorithm could certainly improve the feature.

Another improvement could be to minimize the *types* of content GoGuardian searches through. As we note above, flagging terms in source code creates many unnecessary flags—if websites include "harmful" or "explicit" keywords in source code, or in links, our research shows that these sites are likely *not* themselves explicit or harmful. Generally, these are mistakes.

Expanding the built-in white list would also help. There is little reason to flag Wikipedia, in addition to .edu and .gov sites. Any domain run by common educational companies that schools intentionally use should also be included in a white list.

The best way to fix these issues, of course, is to end this type of flagging entirely.

At the very least, GoGuardian and other companies selling similar student-monitoring snake oil should stop hiding the ball from teachers, administrators, parents and students, and start offering real information about what their products do. To improve public awareness of the software's effectiveness, GoGuardian should make all keyword lists public, and update these lists as the keywords change. (We asked GoGuardian for

their keyword list, but were told that only customers have access to the list of terms. You can find a list of the keywords that were included in the results of our records requests here.) As requested by senators. Warren and Markey, the company should examine the impact of its algorithms on protected classes of students and transparently share the results. As part of this, it should also make available regular, recent, real-life samples of flagged, anonymized data, so anyone can see how effective this software is without having to file FOIA requests. The company should also ensure this data includes clear results of Beacon flags and interventions. Lastly, it should allow third-party auditors to review the software's effectiveness regularly, and publish the results of these audits.

To improve privacy issues, the company should institute default data minimization and deletion requirements. Student data should not be available to anyone with access to the product months or years after it was flagged and collected by the software. Additionally, the company should never * * send data directly to parents or law enforcement that has not been reviewed by the school. GoGuardian should also offer students more clarity on what type of data it collects by allowing anyone that the software flags to see what has been flagged for them and why.

## What Government Can Do to Protect Students from Dangerous Student Monitoring Tools

These issues are primarily GoGuardian's fault and responsibility. But there is a role for the government here, as well.

In our correspondence with GoGuardian, the company repeatedly told us that its keyword list "includes only terms directly linked to the school's requirements under CIPA regarding explicit and harmful content." CIPA, the Children's Internet Protection Act, passed by Congress in 2000, essentially has three requirements: Schools and public libraries that receive federal funding must monitor minors' "online activities," must restrict minors' internet access to "visual depictions" that are "obscene, child pornography, or harmful to minors," and must educate minors about appropriate online behavior.

GoGuardian seems to be interpreting CIPA to require far more than it actually does. One of GoGuardian's main features is to block and filter online pages. CIPA does not require the blocking of keywords or text, and the company's claim that its tools are simply helping schools comply with federal law is confused at best and disingenuous at worst.

As for CIPA's requirement that schools engage in "monitoring the online activities of minors," the FCC must provide clarification as to what the provision does require. FCC's guidance explicitly states that "CIPA does not require the tracking of Internet use by minors or adults." This is true for adults, but does not appear to be true for minors based on the plain text of the statute. The FCC should explicitly state that CIPA does not require any filtering or blocking of keywords or text.

This is why we suggest, alongside Sen. Warren and Sen. Markey, that the FCC issue new guidance related to compliance with CIPA and provide clarification regarding exactly how schools should interpret the requirement to monitor the "online activities" of students. In our view, the tracking of students is not required by CIPA, making much of the functionality of student "monitoring" software not only dangerous but unnecessary. It appears that Congress simply wanted minors to be "monitored" to the extent that obscene and harmful "visual depictions" will be blocked when a student navigates to such images, but further monitoring is not necessary. Regardless, GoGuardian's flagging certainly goes beyond CIPA's requirements. This is not to mention that GoGuardian and other tools like it create practical hindrances for students. According to CDT's research, 71 percent of students whose schools use filtering and blocking software agree that it is sometimes hard to complete school assignments because of these tools.

Schools and libraries are required by CIPA to provide reasonable notice and hold at least one public hearing or meeting when adopting an internet safety policy. Additional hearings or meetings are not required, even if the policy is amended, unless required by state or local rules or by the policy itself, so there is generally no requirement for schools to have additional meetings after adopting new software. CIPA could be amended to require additional notice and meetings before any new policies are adopted, which should include new surveillance or filtering tools.

Senators Warren and Markey also recommend that the U.S. Department of Education require local education agencies to track the potential impacts of these tools on students in protected classes, including data on the use of student activity monitoring tools for disciplinary purposes and other disparate effects. We agree.

Lastly, we have concern that companies like GoGuardian may pivot as the flaws in their tools become more commonly known. They may describe their software as updated, better, having smarter features, or even using AI to determine what should be blocked (many companies have already begun describing their tools as "AI" when they should not). All of this may be true—but government agencies must investigate these claims, and GoGuardian and similar companies must allow them to do so, as we explained above.

## Schools Can Do Better for Students

In the future, we will look back on this moment in time as one when school administrators, hamstrung by limited resources, tried to combat increasing numbers of school shootings and a growing mental health crisis in young people by relying on faulty, inaccurate surveillance software. There is no evidence—except unsupported claims peddled by the companies themselves—that this sort of student monitoring software benefits young people, and there is significant evidence that it harms them. We hope that school administrators begin to recognize that these tools not only fail at what they are supposed to do, but in practice, do something much worse: endanger the very students they are meant to protect.

School officials concerned about student access to harmful content online have many options for filtering tools that don't also track young people's every online action. No filtering system will be perfect; as you can see from this research, if even a third of what is flagged is blocked, it would prevent students from accessing important health information, historical and educational content, literature, and more. But regardless of what tools they use, school administrators should follow some standard practices to ensure the safety of students.

First, schools should train teachers, students, and parents on how this software works, and what it does. (CDT's reporting shows that only 69% of teachers have been trained on this software, and about a third of parents don't know what happens when results occur outside of school hours.) Schools should also make it clear who has access to the data collected by these tools, and set strict data retention limits. They should ensure that the software is not tracking students outside of school property or school hours (40% of teachers report that the school monitors students' personal devices.).

*Under no circumstance* should schools send law enforcement, including student resource officers, information from this software. Lastly, before adopting an internet safety policy, CIPA mandates schools and libraries to provide reasonable notice and hold at least one public hearing or meeting to address the proposal. This should include describing any monitoring or filtering software. Unfortunately, additional hearings or meetings are not required, even if the policy is amended, unless required by state or local rules or by the policy itself, so there is generally no reason for schools to have additional meetings after adopting new software.

## What Parents and Students Can Do

Parents and students working together can make a big impact on a school. Parents should generally be allowed to opt their children out of unnecessary tracking—CDT's reporting shows that most parents (57% want a choice in whether their child is monitored by the

School—but this is often not an option. With the knowledge that this software is so often inaccurate, parents and students should better be able to advocate for changes in a school's use of these tools, such as requiring opt-in consent. Armed with information from EFF, CDT, and other organizations, you *can* make a difference.

If you want to know more about what tools your school uses, ask them! You may also file public records requests, like we did, for your specific school to learn more about how the software works. If you're interested in doing this, email jason@eff.org to get more information on how to file these requests.

You can also help set better standards for schools that are considering (or are already using) various types of surveillance technology or software. Some groups like the ACLU have begun proposing bills that would require schools to be more transparent and thoughtful about surveillance purchases: the Student Surveillance Technology

Acquisition Standards Act is an example. This template would create better standards for schools considering surveillance technology, requiring efficacy and proof before implementation.

Though it refers to a different technology from student monitoring software, EFF supported a California bill to protect students from invasive proctoring software, the Student Test Taker Privacy Protection Act. This law "prohibits a business providing proctoring services in an educational setting from collecting, retaining, using, or disclosing personal information except to the extent necessary to provide those proctoring services." It may provide some ideas for similar legislation to protect young people from student monitoring software, or other surveillance and tracking.

Lastly, you should reach out to us if your online activity has been flagged, or blocked, inaccurately by a student monitoring tool. Email jason@eff.org to let us know more.

# Conclusion

Schools should be safe places for students, but they must also be places where students feel safe exploring ideas. Student monitoring software not only hinders that exploration, but endangers those who are already vulnerable. Student monitoring software creates a prison-like atmosphere that is opposed to a learning environment, and is particularly dangerous for LGBTQ+ people and historically oppressed groups.

Students are better off when we don't use surveillance technology to track their every move, either in person or on the internet. Studies have shown that young people don't want to be surveilled—a study by EDRi showed, for examplethat kids themselves don't want government scanning of their messages, even if it's to prevent child abuse.

Student monitoring software is just one of many types of disciplinary technologies, which typically show up in the areas of life where power imbalances are the norm: in our workplaces, our homes, and of course, our schools. This makes it difficult for the people they impact to fight back and protect themselves. Even with this report, the work done by the Center for Democracy and Technology, and senators Warren and Markey, we know it will be an uphill battle to protect students from surveillance software. Still, it is our hope that this research will help people in positions of authority such as government officials and school administrators, as well as parents and students, to push the companies that make this software to improve, or to abandon their use entirely.

There are *dozens* of different student monitoring software products. While we sometimes make general arguments about the software based on our research, most of our results apply specifically to GoGuardian and its features. However, if other tools that filter, block, or flag material operate similarly to GoGuardian's flagging behavior, then an entire industry is propped up by false flags. We expect that's the case; a 2016 New York Times piece asking students if the filters at their schools were too restrictive got some of the most responses to anything ever written by the paper.

If you have encountered issues with student monitoring software, or have successfully campaigned to change your school's use of it, we'd love to hear from you. Please email us at info@eff.org. Also, if you have further questions about the software, please reach out to us.