Challenges of Child Internet Protection in the Social Media Age

Real-time AI's crucial role in ensuring internet safety

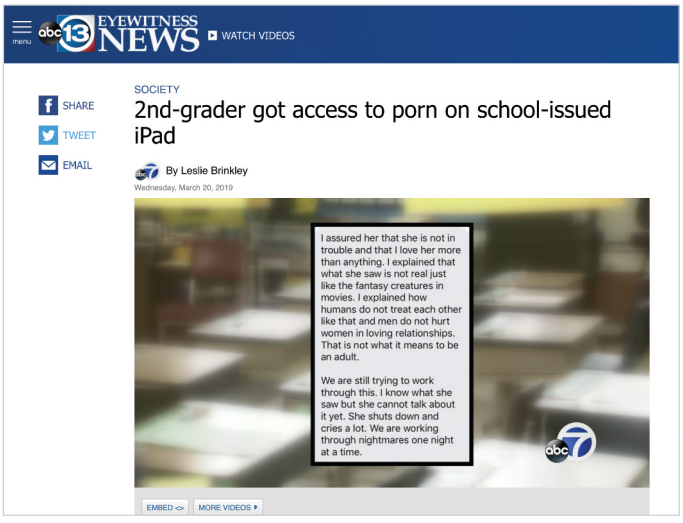
Previous to Deledao, child internet protection solutions focused on static content filtering. Web domains are statically classified and stored in databases owned by the vendors. When a URL is visited, its domain or host name is compared against the database. If it is known to be harmful, access will not be granted.

However, the modern age of social media and user-generated content changed everything.

Social media gained huge popularity around the world because of its viral nature and it is particularly popular among children. Children can keep up with world news, follow their favorite music or sports stars, and form private circles among their friends in social media apps. However, there is an abundance of inappropriate content on social media as well. This presents some new challenges.

- Web domains can no longer be statically cataloged. For instance, there's a wealth of content on YouTube. According to an interview on CBS's 60 Minutes, YouTube's CEO revealed that a staggering 500 hours of new content is uploaded to YouTube every minute. Some content is appropriate while some content can be harmful for children. With legacy solutions, it's an all-or-nothing approach. A social media site is either completely blocked or left wide open. Sites such as YouTube are more likely to be left open than closed in most schools, unknowingly subjecting children to inappropriate content.

  Example – 2nd-grader got access to porn on school-issued iPad



The proliferation of social media and user-generated content has exposed legacy products' reliance on antiquated domain and keyword blocking filtering practices.

- Social media sites almost always use the encrypted HTTPS protocol during data exchange. This coincides with an industry-wide trend to migrate to HTTP/2, which negotiates connections using HTTPS by default. HTTPS is designed to prevent other devices from eavesdropping on the information being transferred. How to decrypt HTTPS traffic for inspection poses a great challenge to legacy solutions.
- In addition to being exposed to harmful content, children are also subject to bullying by their peers on social media. Many children do not know how to handle the pressure, which sometimes results in attempts of self-harm. Troubled children may also share violent tendencies on their social media accounts, which may be early signs of school shooting incidents. How to detect such incidents and sentiments becomes very important. It is equally important to provide accurate visibility to school administrators and parents so that these children can get the help and support they need.

Woefully inadequate in filtering today's internet, legacy solutions often fall into two categories: DNS filtering or proxy-level filtering.

Legacy Solution: DNS Filtering

DNS-based solutions, such as OpenDNS, are the simplest. DNS is the protocol that maps a host name to its IP address on the internet. To deploy a DNS-based solution, the school sets up the computers to use the vendor's DNS servers. Before a webpage is loaded, the browser would query the DNS server with the host name in the URL to find out the IP address of the web server. The filtering DNS server would then compare the provided host name against its category database. If it falls into one of the blocked categories, a different IP address instead of the real web server will be returned to the browser. This basically redirects the browser to a web server hosted by the vendor that will display a page saying the original site is blocked.

This type of solution is very limited in that it makes decisions on whether to allow or block based on the host name part of the URLs only. This makes it ineffective against social media sites (e.g. unblocked Google Sites that host games on its webpages). No fine-grained control at the URL level or webpage content inspection is possible.

Legacy Solution: Proxy-Level Filtering

Proxy-level filters are deployed at network gateways to intercept web traffic and inspect it. They are an improvement over DNS-based solutions in that they can see all HTTP content being transferred to the browser. In addition to the host names, they can also see the full URLs, HTTP requests and response details.

However, its solution to HTTPS traffic is very cumbersome. In order to examine such traffic, the proxy-level filter has to perform a man-in-the-middle attack by presenting a fake certificate to the client browser. This often involves deploying a new root CA certificate to all browsers in the organization, which is a complex process. What complicates the situation more is that many browsers now support certificate pinning and would simply reject these fake certificates, making it impossible for the proxy-level filter to inspect traffic. Also, the non-browser apps, which might not recognize the proxy-level filter's root CA certificates, will also reject them and stop functioning. The visibility into the overall browser traffic is rapidly declining.

Even if the proxy-level filter can see HTTPS traffic, it's virtually impossible to perform meaningful content analysis on every webpage for performance reasons. Suppose each analysis will take 5ms. If there are 500 students browsing the web simultaneously and every webpage is inspected, it will cause a latency of 5ms * 500 / 4 = 750ms with a 4-core CPU. This slowdown will surely be noticeable by the users. This is why proxy-level filters perform very little scanning on actual web content being transferred, but instead rely on the pre-generated domain category database.

The dynamic nature of modern webpages makes it even more difficult to inspect content. For instance, the main HTML files of many websites (such as www.cnn.com) almost exclusively contain JavaScript code, which in turn loads additional content from the web server in separate HTTP requests to render in the browser. The pages of websites like www.facebook.com have implemented infinite scrolling. As a user scrolls down the page, additional content is loaded from the server to display in the browser, which makes the page content appear to be infinite. This means the HTTP requests that the proxy servers are seeing are all fragments of the same webpage. It's virtually impossible for them to track and assemble all the fragments to try to figure out what is rendered in the browsers.

The usage reports generated by proxy-level filters are based on the number of requests to each web server seen by the proxies. For instance, if it sees 5 HTTP requests to www.cnn.com from a certain user, it would give www.cnn.com a weight of 5 when calculating the sites the user has spent the most time on. This would heavily skew the results toward webpages that have a lot of embedded objects (e.g. images, scripts, etc.) over those that have simple HTML content. The correct measurement of such reports should be how much time a user spends viewing a page, not how many objects are loaded on each page.

Almost all of the above problems stem from the fact that what the proxies see are not what the users see in the browsers. Further, proxy-level filters are cumbersome to deploy, may bear additional upfront costs because of the physical or virtual appliances involved, and are single points of failure on the network. Recently, there are cloud-based proxy solutions for which no appliances need to be deployed on premise. However, some of them may be even more difficult to deploy than on-premise proxy filters and remain single points of failure for all web traffic.

The Deledao Advantage: Network Filtering Experts

Though a relative newcomer to the market, Deledao is certainly not new to the filtering industry. Deledao's co-founder and COO, Shuang Ji, co-invented technology used in proxy scanning and filtering in the late 1990s. This technology serves as the foundation of most proxy-level filter products today.

As a father of a school-aged child himself, Shuang sees firsthand how outdated content filtering fails to respond to the modern web. Alongside fellow Silicon Valley veterans, the Deledao team developed the first ever real-time artificial intelligence content filter for schools.

Our founders possess decades of experience in network filtering. They are keenly aware of how AI will better meet the filtering needs of today's dynamic internet.

The Deledao Advantage: AI Protection at the Browser Level

Deledao uses an innovative approach to solve the challenges of a modern internet – Deledao delivers AI technology to filter at the endpoint.

Instead of seeing individual HTTP objects, Deledao sees how each object displays within the browser. Then, Deledao's proprietary artificial intelligence engines analyze text, image and video content in real time. By using AI to filter at the browser level, Deledao can see exactly what the users see and can analyze each webpage like a human would. The challenges faced by current solutions no longer apply:

- SSL decryption for HTTPS traffic is not a problem. When the traffic reaches the browser, it is already decrypted, then it's filtered by Deledao.
- Dynamic page content generated by script code (e.g. pages with infinite scrolling) or multiple HTTP requests don't pose problems, either. The Deledao solution is not sensitive to how page content loads from the network, but instead focuses on the rendered content on the UI. Therefore, it can reliably extract the text and images being displayed in the browser for analysis.
- The browser knows exactly when the user is navigating to a new webpage and can record the time the user has spent on each page with an accuracy of milliseconds.
- Reports show page titles and page content vs only showing URLs for you to decipher web activity.
- The Chrome extension and browser deployment options are VPN-proof.
- Deledao is 100% cloud-based software. There is no hardware to deploy or maintain on premise.

Deledao is a 100% cloud-based content filter software that uses AI to filter at the browser level.

The Deledao Advantage: AI Filter vs AI Database

The distinction between using AI to filter in real time vs using an AI-compiled database is crucial.

The challenges faced by legacy solutions are not solved by using an AI database. An AI database implies that AI or machine learning technology is used to pre-categorize websites and keywords. How would it be able to handle webpages with dynamic content, such as video or pages with infinite scrolling?

Whether it's an endpoint solution or not, AI filtering at the proxy level uses an AI database. It's not technically feasible to have an AI filter at the proxy level due to the amount of data that passes through this bottleneck. Claims that AI is being used at that level probably refer to an AI database.

By employing AI filtering technologies in real-time, Deledao is able to analyze every webpage, instead of relying on a domain category database. For instance, it can easily read the whole text and tell the difference between Google search results for "adult education" vs "adult movies". With legacy filter products, one would need to either block the entire google.com domain or the keyword "adult" in search keywords, either of which would result in considerable undesired side-effects.

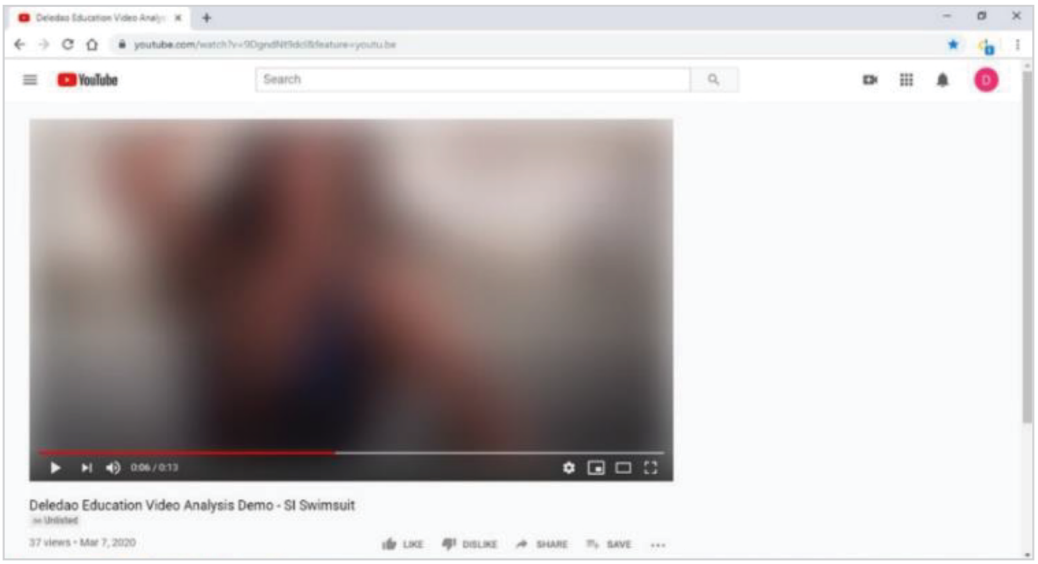
It's worth finding out if your current solution uses AI in real time for every webpage vs using an AI database.

Deledao uses AI to filter in real time instead of relying on AI to compile a database beforehand.

The Deledao Advantage: Innovative Image and Video Filtering

YouTube and other online video platforms pose problems for legacy solutions. Typical solutions only allow blocking by channel or category. Unique to Deledao, our proprietary AI filtering technology blurs and mutes inappropriate video content in real time.

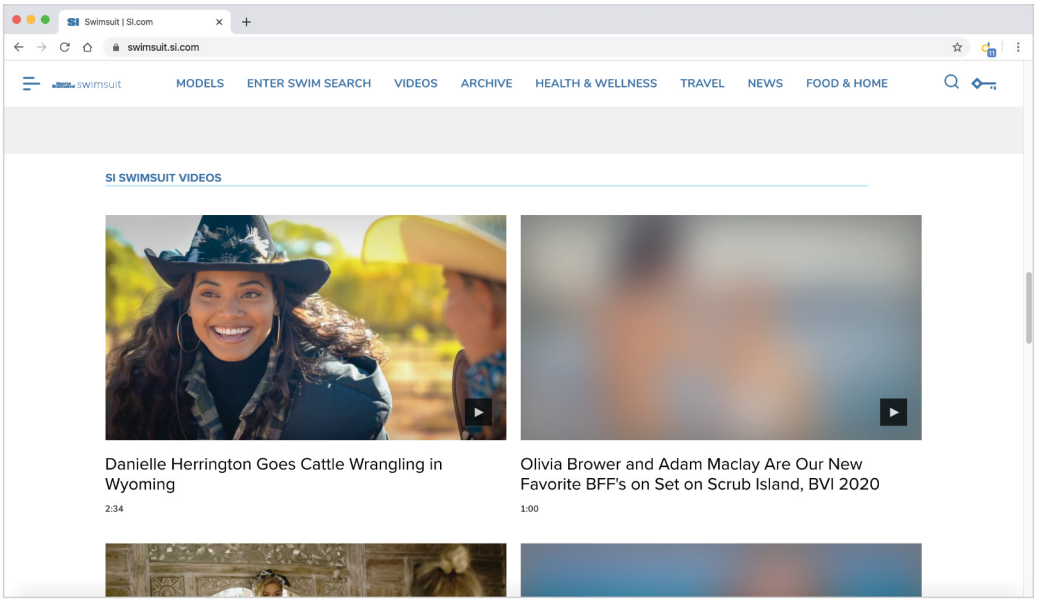
Example – Filtering Sports Illustrated swimsuit videos on YouTube



We also provide IT admins with the option to disable comments and remove sidebar video suggestions on YouTube.

When it comes to image filtering, our proprietary AI technology also works in real time. Whenever an inappropriate image renders in the browser, the image will blur on the webpage.

Example – Filtering Sports Illustrated swimsuit images on the website



Deledao takes the claim "real-time filtering" seriously. View any online image or play any online video and Deledao will blur the image and blur and mute the inappropriate video scenes in real time.

The Deledao Advantage: Device and OS Agnostic

Why should your content filter work on some devices but not others?

Our AI filtering technology is compatible with Chromebooks, Windows, macOS, and iOS devices.

Unlike many content filters on the market, Deledao works seamlessly on iPads. Other solutions struggle with filtering on iPads due to decisions that Apple made as a company. Many content filters utilize a combination of static Block Lists installed on the school device and PAC (Proxy Auto Config) files to direct some traffic to a cloud proxy for filtering. In this case, all the problems with proxy-level filtering still apply. Because Deledao's technology works at the browser level, we can bypass these issues.

Deledao works on iPads, Chromebooks, Windows, Mac OS, and tablets with deployment often taking less than 30 minutes.

Conclusion:

In summary, legacy filtering technologies are simply unable to effectively manage today's internet for on-campus and off-campus students. Deledao is a true AI filter that screens through each webpage in real time to analyze words, images, and videos in context. No other filter product on the market provides this level of sophisticated content filtering.