

# Appropriate Filtering for Education settings



May 2023

## Filtering Provider Checklist Responses

Schools (and registered childcare providers) in England and Wales are required “to ensure children are safe from terrorist and extremist material when accessing the internet in school, including by establishing appropriate levels of filtering”. Furthermore, it expects that they “assess the risk of [their] children being drawn into terrorism, including support for extremist ideas that are part of terrorist ideology”. There are a number of self review systems (eg [www.360safe.org.uk](http://www.360safe.org.uk)) that will support a school in assessing their wider online safety policy and practice.

The Department for Education’s statutory guidance ‘Keeping Children Safe in Education’ obliges schools and colleges in England to “ensure appropriate filters and appropriate monitoring systems are in place and regularly review their effectiveness” and they “should be doing all that they reasonably can to limit children’s exposure to [Content, Contact, Conduct, Contract] risks from the school’s or college’s IT system” however, schools will need to “be careful that “over blocking” does not lead to unreasonable restrictions as to what children can be taught with regards to online teaching and safeguarding.”

By completing all fields and returning to UK Safer Internet Centre ([enquiries@saferinternet.org.uk](mailto:enquiries@saferinternet.org.uk)), the aim of this document is to help filtering providers to illustrate to education settings (including Early years, schools and FE) how their particular technology system(s) meets the national defined ‘appropriate filtering standards. Fully completed forms will be hosted on the UK Safer Internet Centre website alongside the definitions

It is important to recognise that no filtering systems can be 100% effective and need to be supported with good teaching and learning practice and effective supervision.

Company / Organisation	Diladele B.V.
Address	Ko Donckerlaan, 26, 1187TE, Amstelveen, The Netherlands
Contact details	Rafael Akchurin, <a href="mailto:rafael.akchurin@diladele.com">rafael.akchurin@diladele.com</a>
Filtering System	Web Safety ( <a href="https://www.diladele.com">https://www.diladele.com</a> ), version 8.5 and above.
Date of assessment	22 June, 2023

## System Rating response

Where a supplier is able to confirm that their service fully meets the issue identified in a specific checklist the appropriate self-certification colour for that question is GREEN.	
Where a supplier is not able to confirm that their service fully meets the issue identified in a specific checklist question the appropriate self-certification colour for that question is AMBER.	

## Illegal Online Content

Filtering providers should ensure that access to illegal content is blocked, specifically that the filtering providers:

Aspect	Rating	Explanation
<ul style="list-style-type: none"> <li>Are IWF members</li> </ul>		Member since 1 December, 2016. See <a href="https://www.iwf.org.uk/membership/our-members/diladele-by/">https://www.iwf.org.uk/membership/our-members/diladele-by/</a>
<ul style="list-style-type: none"> <li>and block access to illegal Child Abuse Images (by actively implementing the IWF URL list)</li> </ul>		IWF block list is integrated into Web Safety.
<ul style="list-style-type: none"> <li>Integrate the 'the police assessed list of unlawful terrorist content, produced on behalf of the Home Office'</li> </ul>		The block list produced by Home Office UK (CTIRU block list) is integrated into Web Safety.
<ul style="list-style-type: none"> <li>Confirm that filters for illegal content cannot be disabled by the school</li> </ul>		It is not possible to disable a filter from outside of the deployed application.

## Inappropriate Online Content

Recognising that no filter can guarantee to be 100% effective, providers should both confirm, and describe how, their system manages the following content

Content	Explanatory notes – Content that:	Rating	Explanation
Discrimination	Promotes the unjust or prejudicial treatment of people on the grounds of race, religion, age, or sex.		Web Safety provides "Hate / Discrimination / Violence" category that includes sites promoting racial hatred, violence and homophobia. Blocking is performed based on URL database with periodic updates.
Drugs / Substance abuse	displays or promotes the illegal use of drugs or substances		Web Safety provides "Drugs" category that contains pages related to drugs, narcotics and other substances. Blocking happens not only based on URLs but also can be applied dynamically on any page. Web Safety also has "Tobacco and Alcohol" category that can be blocked in more restrictive environments.
Extremism	promotes terrorism and terrorist ideologies, violence or intolerance		Such sites are blocked by CTIRU list. Web Safety does not have a separate category and relies on CTIRU list for blocking access to such sites.
Gambling	Enables gambling		Web pages are dynamically assigned the "Gambling" category using Artificial

			Intelligence text analyser module. Site triggering positive response from that module are blocked.
Malware / Hacking	promotes the compromising of systems including anonymous browsing and other filter bypass tools as well as sites hosting malicious content		Sites assigned the “Hacking, Cracking and Illegal Content” category are blocked. Blocking is performed based on URL database with periodic updates.
Pornography	displays sexual acts or explicit images		Web Safety provides “Nudity / Pornography” category as well as ability to dynamically scan downloaded web pages for any adult content. Using latest machine learning techniques, Web Safety is capable of blocking any reference to adult materials even on sites general in nature (e.g., searches for adult content on Google, Bing, Yahoo, YouTube).
Piracy and copyright theft	includes illegal provision of copyrighted material		Sites assigned the “Hacking, Cracking and Illegal Content” category are blocked. Blocking is performed based on URL database with periodic updates.
Self Harm	promotes or displays deliberate self harm (including suicide and eating disorders)		Sites assigned the “Suicide and self-harm” category are blocked. Blocking is performed based on URL database with periodic updates as well as dynamic deep content categorization.
Violence	Displays or promotes the use of physical force intended to hurt or kill		Web Safety includes “Hate / Discrimination / Violence” category. Blocking is performed based on URL database with periodic updates. We also support dynamic “Weapons” category, that allows deep content analysis and blocking of pages of this category on the go.

This list should not be considered an exhaustive list. Please outline how the system manages this content and many other aspects

Web Safety is different from most other web filtering solutions in that it “looks into” the traffic that is being filtered. It uses latest advancements in Machine Learning and Natural Language Processing to provide efficient zero-day protection. This allows blocking of questionable material even on sites general in nature.

Web Safety also allows performing blocking using black lists of words that can be easily adjusted by the administrator to fine tune what exactly gets blocked without knowledge of data science. This approach proves to be very effective in filtering.

Administrators are also able to manually re-categorize sites which are not yet known to the system and thus block access to such sites. The list of re-categorized sites can be shared with Diladele that in turn leads to inclusion of these sites into main categorization database. After internal verification, access to re-categorized sites is blocked for all users of the application.

Our dynamic categorization modules can be enabled to dynamically categorize and block unseen web pages that are not contained in our database. The list of dynamic categories is growing with each release of our application and at the moment of this writing includes the following categories: "Adult", "Alcohol and Tobacco", "Dating", "Drugs", "Gambling", "Games", "Pornography/Nudity", "Suicide and self-harm" and "Weapons" (the above categories are currently supported only in English).

Regarding the duration and extent of logfile (Internet history) data retention, providers should outline their retention policy, specifically including the extent to the identification of individuals and the duration to which all data is retained.

Administrators of the system can configure log retention policy up to several years if required. The length of log history is only limited by the available hard disk storage for logs. Traffic Monitoring module periodically builds reports of browsing activities for the users of the system. These reports can be viewed in the Admin UI or sent by e-mail as configured by the system administrator.

It is also possible to authenticate users and/or anonymize logs to minimize privacy violations. Some categories of sites (banking, financial, government, health) are automatically excluded from deep content inspection and filtering to minimize exposure of private information from browsing users. Admin UI can be used to fine tune the excluded sites, categories and authenticated users.

Providers should be clear how their system does not over block access so it does not lead to unreasonable restrictions

If parts of a site are blocked incorrectly it is always possible to exclude them or the site as a whole from web filtering. Access then becomes unrestricted and no filtering takes place.

If a site is incorrectly categorized, administrator is able to re-categorize it. New categories of a site may be automatically uploaded to our servers and after manual verification categories of that site can be adjusted. After definition database updates all users of the application start to categorize that site correctly.

We thoroughly investigate all reported false positives and negatives of dynamic categorization as well and constantly improve our machine learning algorithms.

## Filtering System Features

How does the filtering system meet the following principles:

Principle	Rating	Explanation
<ul style="list-style-type: none"> <li>Context appropriate differentiated filtering, based on age, vulnerability and risk of harm – also includes the ability to vary filtering strength appropriate for staff</li> </ul>		<p>Web safety can apply different web filtering policies for different groups of users. It is possible to identify users by IP, user name or Active Directory security group memberships.</p>
<ul style="list-style-type: none"> <li>Circumvention – the extent and ability to identify and manage technologies and techniques used to circumvent the system, specifically VPN, proxy services and DNS over HTTPS.</li> </ul>		<p>Web Safety blocks access to large list of known anonymous proxies and VPN sites. Traffic being sent to all yet unknown proxies is decrypted and filtered, preventing access to all prohibited contents automatically.</p>
<ul style="list-style-type: none"> <li>Control – has the ability and ease of use that allows schools to control the filter themselves to permit or deny access to specific content. Any changes to the filter system are logged enabling an audit trail that ensure transparency and that individuals are not able to make unilateral changes</li> </ul>		<p>Web Safety is deployed on-site by the school administrator himself and all settings of the application are managed by the administrator. Access to any content can be denied/permitted by the administrator.</p>
<ul style="list-style-type: none"> <li>Contextual Content Filters – in addition to URL or IP based filtering, the extent to which (http and https) content is analysed as it is streamed to the user and blocked, this would include AI generated content. For example, being able to contextually analyse text on a page and dynamically filter.</li> </ul>		<p>The application is able to inspect HTTPS sessions and analyse the actual contents of the downloaded pages. The dynamic categorization module, URL heuristics and page words analysis are used then to scan and possibly block the accessed page.</p>
<ul style="list-style-type: none"> <li>Filtering Policy – the filtering provider publishes a rationale that details their approach to filtering with classification and categorisation as well as over blocking</li> </ul>		<p>Web filtering policies and how these are to be managed are described at our online admin guides for each version. See <a href="https://docs.diladele.com">https://docs.diladele.com</a></p>
<ul style="list-style-type: none"> <li>Group / Multi-site Management – the ability for deployment of central policy and central oversight or dashboard</li> </ul>		<p>Web Safety supports clusters of machines with a centralized management and configuration.</p>
<ul style="list-style-type: none"> <li>Identification - the filtering system should have the ability to identify users</li> </ul>		<p>Web Safety is capable of user authentication using IP/MAC labelling, IP address, range, subnets. Can be easily integrated with Microsoft Active Directory and can</p>

		identify users by user name/password generated by the administrator.
<ul style="list-style-type: none"> <li>Mobile and App content – mobile and app content is often delivered in entirely different mechanisms from that delivered through a traditional web browser. To what extent does the filter system block inappropriate content via mobile and app technologies (beyond typical web browser delivered content). Providers should be clear about the capacity of their filtering system to manage content on mobile and web apps</li> </ul>		<p>Web Safety is able to decrypt HTTPS connections typically used by mobile applications.</p> <p>Decrypted connections are filtered as usual. Mobile applications which do not support decrypted connections are blocked automatically (stop functioning themselves because of the decrypted connection).</p>
<ul style="list-style-type: none"> <li>Multiple language support – the ability for the system to manage relevant languages</li> </ul>		<p>Deep content inspection module is based on UTF-8 Unicode matching – it means it can find and block any content in any language (because internally all languages are recoded into uniform Unicode representation before actual matching). We have proved deployments that successfully block access to explicit content in Hebrew for example.</p> <p>Nevertheless, out of the box the solution is targeted to Western languages blocking mostly.</p>
<ul style="list-style-type: none"> <li>Network level - filtering should be applied at 'network level' i.e., not reliant on any software on user devices whilst at school (recognising that device configuration/software may be required for filtering beyond the school infrastructure)</li> </ul>		Web Safety is deployed as either explicit or transparent proxy within the school network.
<ul style="list-style-type: none"> <li>Remote devices – with many children and staff working remotely, the ability for school owned devices to receive the same or equivalent filtering to that provided in school</li> </ul>		Web Safety is installed at school on premises, so the end user device *must* be setup to route traffic to the application, for example using always on VPN.
<ul style="list-style-type: none"> <li>Reporting mechanism – the ability to report inappropriate content for access or blocking</li> </ul>		Web Safety contains approximately 25 reports for various user activities. Additional reports can be

		manually built by the administrator
<ul style="list-style-type: none"> <li>• Reports – the system offers clear historical information on the websites users have accessed or attempted to access</li> </ul>		Web Safety is able to store information of every visited URL. The length of log history is configurable by the administrator. Logs can be anonymized if required.
<ul style="list-style-type: none"> <li>• Safe Search – the ability to enforce 'safe search' when using search engines</li> </ul>		Safe Search is supported on Google, Bing and Yahoo. Even on unknown search engines all inappropriate content can be blocked due to the deep content inspection modules.

Filtering systems are only ever a tool in helping to safeguard children when online and schools have an obligation to *“consider how children may be taught about safeguarding, including online, through teaching and learning opportunities, as part of providing a broad and balanced curriculum”*.<sup>1</sup>

Please note below opportunities to support schools (and other settings) in this regard

Reports in Web Safety can be customized by the browsing user name. This allows the teacher to see what sites the user browsed, what searches on Google/Bing/Yahoo did and what videos were watched on YouTube.

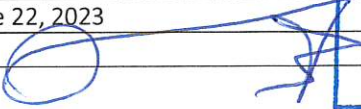
---

<sup>1</sup> <https://www.gov.uk/government/publications/keeping-children-safe-in-education--2>

### PROVIDER SELF-CERTIFICATION DECLARATION

In order that schools can be confident regarding the accuracy of the self-certification statements, the supplier confirms:

- that their self-certification responses have been fully and accurately completed by a person or persons who are competent in the relevant fields
- that they will update their self-certification responses promptly when changes to the service or its terms and conditions would result in their existing compliance statement no longer being accurate or complete
- that they will provide any additional information or clarification sought as part of the self-certification process
- that if at any time, the UK Safer Internet Centre is of the view that any element or elements of a provider's self-certification responses require independent verification, they will agree to that independent verification, supply all necessary clarification requested, meet the associated verification costs, or withdraw their self-certification submission.

Name	Rafael Akchurin
Position	CEO
Date	June 22, 2023
Signature	

**Diladele B.V.**  
Ko Donckerlaan 26  
1187TE, Amstelveen, NL  
mail@diladele.com