Appropriate Filtering for Education settings

May 2025

Filtering Provider Checklist Reponses

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) 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), 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	Exa Networks Ltd
Address	100 Bolton Road, Bradford, BD1 4DE
Contact details	0345 145 1234
Filtering System	SurfProtect
Date of assessment	12/09/2025

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
● Are IWF members	🟩	Exa has been an IWF member since 2005. https://www.iwf.org.uk/membership/our-members/exa-networks/
● and block access to illegal Child Abuse Images (by actively implementing the IWF URL list), including frequency of URL list update	🟩	We block the IWF list by default, with no ability to disable it. We query the IWF API multiple times a day for updates to their lists.
● Integrate the 'the police assessed list of unlawful terrorist content, produced on behalf of the Home Office'	🟩	We block the CTIRU list by default, with no ability to disable it.
● Confirm that filters for illegal content cannot be disabled by anyone at the school (including any system administrator).	🟩	It is impossible to disable or bypass the above lists through SurfProtect. This also applies to: ● PIPCU

Additional illegal content blocked by SurfProtect:

- **PIPCU**: SurfProtect implements the City of London Police's Infringing Website List (IWL) from the Police Intellectual Property Crime Unit (PIPCU). This list is used to block websites that distribute or promote copyrighted material without permission, reducing digital piracy and associated criminal activity.
- **PDNS**: SurfProtect enforces the PDNS service provided by the NCSC (National Cyber Security Center) , another layer of security to protect schools from malware.

Describing how, their system manages the following illegal content

Note: Exa defines a set of illegal content categories that are automatically blocked and cannot be removed. Any content identified within these categories is immediately restricted and remains inaccessible. Additionally, our real-time alerts system will send email notifications to a school's nominated contacts when their users send requests for sites that are in these categories.

Content	Explanatory notes – Content that:	Rating	Explanation
child sexual abuse	Content that depicts or promotes sexual abuse or exploitation of children, which is strictly prohibited and subject to severe legal penalties.	🟩	We address this, CSAM, in the illegal online content section above. This falls under the IWF list which is always blocked and impossible to remove.
controlling or coercive behaviour	Online actions that involve psychological abuse, manipulation, or intimidation to control another	🟩	SurfProtect defines a "Controlling or Coercive Behaviour" category that users cannot opt out of blocking. As such, all websites in

	individual, often occurring in domestic contexts.	🟩	the category are automatically blocked.
extreme sexual violence	Content that graphically depicts acts of severe sexual violence, intended to shock or incite similar behaviour, and is illegal under UK law.	🟩	SurfProtect defines a “Extreme Sexual Violence” category that users cannot opt out of blocking. As such, all websites in the category are automatically blocked.
extreme pornography	Pornographic material portraying acts that threaten a person's life or could result in serious injury, and is deemed obscene and unlawful.	🟩	SurfProtect defines an “Extreme Pornography“ category that users cannot opt out of blocking. As such, all websites in the category are automatically blocked.
fraud	Deceptive practices conducted online with the intent to secure unfair or unlawful financial gain, including phishing and scam activities.	🟩	SurfProtect defines a “Fraud“ category that users cannot opt out of blocking. As such, all websites in the category are automatically blocked.
racially or religiously aggravated public order offences	Content that incites hatred or violence against individuals based on race or religion, undermining public safety and cohesion.	🟩	SurfProtect defines a “Racially or Religiously Aggravated Public Order Offences“ category that users cannot opt out of blocking. As such, all websites in the category are automatically blocked.
inciting violence	Online material that encourages or glorifies acts of violence, posing significant risks to public safety and order.	🟩	SurfProtect defines a “Inciting Violence“ category that users cannot opt out of blocking. As such, all websites in the category are automatically blocked.
illegal immigration and people smuggling	Content that promotes or facilitates unauthorized entry into a country, including services offering illegal transportation or documentation.	🟩	SurfProtect defines a “Illegal Immigration and People Smuggling“ category that users cannot opt out of blocking. As such, all websites in the category are automatically blocked.
promoting or facilitating suicide	Material that encourages or assists individuals in committing suicide, posing serious risks to vulnerable populations.	🟩	SurfProtect defines a “Promoting or Facilitating Suicide“ category that users cannot opt out of blocking. As such, all websites in the category are automatically blocked.
intimate image abuse	The non-consensual sharing of private sexual images or videos, commonly known as "revenge porn," intended to cause distress or harm.	🟩	SurfProtect defines a “Intimate Image Abuse“ category that users cannot opt out of blocking. As such, all websites in the category are automatically blocked.
selling illegal drugs or weapons	Online activities involving the advertisement or sale of prohibited substances or firearms, contravening legal regulations.	🟩	SurfProtect defines a “Selling Illegal Drugs or Weapons“ category that users cannot opt out of blocking. As such, all

		🟩	websites in the category are automatically blocked.
sexual exploitation	Content that involves taking advantage of individuals sexually for personal gain or profit, including trafficking and forced prostitution.	🟩	SurfProtect defines a “Sexual Exploitation“ category that users cannot opt out of blocking. As such, all websites in the category are automatically blocked.
Terrorism	Material that promotes, incites, or instructs on terrorist activities, aiming to radicalise individuals or coordinate acts of terror.	🟩	As mentioned in the above ‘Illegal Online Content’ section, SurfProtect blocks websites on the CTIRU list. Users cannot opt out of this behaviour.

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
Gambling	Enables gambling	🟩	SurfProtect has a 'Gambling' category, which is blocked by default.
Hate speech / Discriminiation	Content that expresses hate or encourages violence towards a person or group based on something such as disability, race, religion, sex, or sexual orientation. Promotes the unjust or prejudicial treatment of people with protected characteristics of the Equality Act 2010	🟩	This content falls under the ‘Intolerance and hate’ category, but will have aspects that fall under ‘Criminal Activity’ or ‘Tasteless & Offensive’ categories. Additionally, Social Media sites are likely to also contain this content, which is covered by the ‘Social Networking’ category. This is by default blocked.
Harmful content	Content that is bullying, abusive or hateful. Content which depicts or encourages serious violence or injury. Content which encourages dangerous stunts and challenges; including the ingestion, inhalation or exposure to harmful substances.	🟩	The content described here is quite broad, but will fall under the ‘Violence’ category. This may also be covered by other categories such as ‘Criminal Activity’ or ‘Intolerance & Hate’. Harmful content, in general, could be found anywhere such as Social media, Video streaming services. Alongside blocking the related category and search-term based filtering, SurfProtect provides additional filtering settings in these areas like: ● Safe-search ● Youtube restricted mode
Malware / Hacking	promotes the compromising of systems including anonymous	🟩	As stated above, in the Illegal Online Content section,

	browsing and other filter bypass tools as well as sites hosting malicious content	🟩	SurfProtect uses PDNS to block access to known malicious websites and cyber-security threats. Additionally, where this content is not blocked by PDNS it will fall under the categories ‘Hacking’, ‘Phishing/Online Fraud’, ‘Spam URLs’ or ‘Spyware’, all of which are blocked by default.
Mis / Dis Information	Promotes or spreads false or misleading information intended to deceive, manipulate, or harm, including content undermining trust in factual information or institutions	🟩	SurfProtect has a ‘Mis/Dis Information’ category, which is blocked by default.
Piracy and copyright theft	includes illegal provision of copyrighted material	🟩	SurfProtect has an ‘Illegal Filesharing’ category that is blocked by default.
Pornography	displays sexual acts or explicit images	🟩	SurfProtect has an ‘Adult / Sexually Explicit’ category that is enabled by default. Additionally, we have an ‘Intimate Apparel / Swimwear’ category that catches the less explicit, but still inappropriate, content.
Self Harm and eating disorders	content that encourages, promotes, or provides instructions for self harm, eating disorders or suicide	🟩	SurfProtect has a 'Suicide & Self harm' category that is blocked by default. Additionally, our alerting system analyses requests which fall into the ‘Suicide & Self harm’ category, or have identified searches related to this category, and generates alerts and emails to notify nominated contacts.
Violence Against Women and Girls (VAWG)	Promotes or glorifies violence, abuse, coercion, or harmful stereotypes targeting women and girls, including content that normalises gender-based violence or perpetuates misogyny.	🟩	The content falls under the ‘Violence’ category, but may be covered by other categories such as ‘Criminal Activity’ or ‘Intolerance & Hate’.

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

SurfProtect’s host classification database has over 68 million entries, which are split into 61 distinct categories.

Alongside this, we have manually-configured settings (rules for how to apply search term filtering or how the host relates to various application controls) for over 150 thousand hosts, reflecting 20 years of experience in web filtering.

Per DfE guidance, customers are able to set their own categories for websites, allowing them to tailor SurfProtect's functionality to the needs of their organisation and filtering requirements.

SurfProtect also applies search term filtering to a number of websites, including search providers like Google, Bing, Brave, DuckDuckGo and websites like Wikipedia and YouTube, preventing users from making inappropriate searches. We offer 10 "sensible default" lists of search terms that customers can use (such as words related to weapons, drugs, and sex), in addition to allowing customers to define their own list of blocked search terms. These lists are available in the 10 most spoken languages in the UK.

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.

SurfProtect Quantum allows customers to view and retrieve 3 months worth of log history from the current date, whereas SurfProtect Quantum+ allows for 12 months.

The identifying information available in the logs depends on how a customer has configured their SurfProtect service, for their own requirements. At minimum, this will be an external IP address, but can also include internal IP addresses and usernames from any supported user identity provider, such Windows Active Directory, Google Workspace or Microsoft Entra ID (formerly known as Azure Active Directory).

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

SurfProtects default filtering setup provides what we regard as sensible defaults;

- Blocking generally harmful material
- Enabling safe search on search engines, YouTube restricted mode
- Blocking inappropriate or harmful search terms.

Currently, by default, we block AI content as we develop more solutions to combat this new landscape we are presented with, making sure we are in the best position to protect our users.

All these filtering settings, apart from what is mentioned in the initial 'Illegal Online Content' section, can be configured to the needs of the customer. Each of your SurfProtect profiles can be configured differently, allowing you to tailor filtering settings to specific cohorts of users - younger and older students, teachers, safeguarding staff, etc.

Filtering System Features

How does the filtering system meet the following principles:

Principle	Rating	Explanation
● Context appropriate differentiated filtering, based on age, vulnerability and risk of harm – also includes the ability to vary filtering strength appropriate for staff	🟩	SurfProtect allows schools to granularly apply appropriate filtering policies to different sections of their user base via its profile system. Profiles allow administrators to divide up requests received by SurfProtect based one of the following: ● External IP - profiles can be associated with one or more WAN IP addresses or CIDR ranges. ● Internal IP - as above, but for LAN IP addresses. ● Username - profiles can be associated with one or more usernames provided by identity services like AD, Google Workspace and Entra ID. ● Groups - profiles can also be associated with the groups in the above services. ● Proxy port - profiles can match on the port a device is configured to connect to SurfProtect with. The above allows an administrator to apply different policies to different cohorts. For example, if all of your year 7 students are in an AD group, you could create a SurfProtect profile matching that group name and configure it to allow

	🟩	appropriate access for that age group.
● Circumvention – the extent and ability to identify and manage technologies and techniques used to circumvent the system, specifically VPN, proxy services, DNS over HTTPS and ECH.	🟩	**VPNs**: Surfprotect categorises VPNs under a dedicated category "Proxies" , which is blocked by default. **Proxies**: For Proxy like services, such as Google Translate, SurfProtect is able to extract and correctly filter the URL that the user is attempting to view. **Web Proxies**: Our transparent interception technology captures web traffic routed through third-party web proxies. Where transparent SurfProtect interception is not available, steps should be taken to prevent users from altering their device settings in a way that could bypass filtering—such as changing proxy configurations. Preventing users from altering the configuration of their devices lies outside the remit of SurfProtect; therefore, we strongly recommend that schools adopt a defence-in-depth approach. In all cases, customers should implement appropriate device controls to enforce the expected proxy configuration and prevent circumvention. **DNS over HTTPS (DoH)**: SurfProtect is not reliant on DNS for filtering and decrypts all web requests sent to it, which enables it to be able to see the destination URL even if DOH is in use.

	🟩	**Encrypted Client Hello (ECH)**: SurfProtect prevents filtered devices from using encrypted client hellos while negotiating TLS, in order to prevent using ECH as a way to evade filtering.
● 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	🟩	As part of onboarding and provisioning SurfProtect, customers are given access to the SurfProtect panel, where they can customise their filtering settings to suit their specific requirements. All actions performed in the panel generate auditable records, allowing customers to view a history of all changes made to their filtering settings. For MATs and resellers, the system provides a hierarchical access model. Once logged in, they can view and manage settings for the schools or customers they are responsible for, while individual customers retain access and control only over their own settings.
● Contextual Content Filters – in addition to URL or IP based filtering, Schools should understand the extent to which (http and https) content is dynamically analysed as it is streamed to the user and blocked. This would include AI or user generated content, for example, being able to contextually analyse text and dynamically filter the content produced (for example ChatGPT). For schools' strategy or policy that allows the use of AI or user generated content, understanding the technical limitations of the system, such as whether it supports real-time filtering, is important.	🟩	SurfProtect's filtering system fetches and analyses newly visited websites to accurately classify and protect against online risks. To ensure a user is not exposed to inappropriate content during the process, access to the site is restricted while classification is ongoing, which is typically less than 20 seconds. A host's classifications are re-checked to ensure they are still accurate to the content of the website. At present, SurfProtect does not make filtering decisions

	🟩	based on intercepted response content. Due to the user-driven and volatile nature of Generative AI, we have taken the stance that AI and AI-related content should be blocked by default. Schools are able to unblock the AI category and keywords if they deem it appropriate to do so - the following DfE publication states AI can be used in a school environment, but that schools must make safety their top priority when determining if they will do so. https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education.
- Deployment – filtering systems can be deployed in a variety (and combination) of ways (eg on device, network level, cloud, DNS). Providers should describe how their systems are deployed alongside any required configurations	🟩	As SurfProtect is primarily a cloud-based filtering solution, there is very little deployment required. Most of the work is configuration to route web traffic from the customers network to SurfProtect in the cloud, using a range of supported methods depending on the customers requirements. **Note**: In all cases, it is assumed that customers implement appropriate controls on their network to prevent users from bypassing proxy configurations. Configuration can, and often will, occur at both the network and device level: **Network level configuration**: The customer's equipment is configured to route web traffic to SurfProtect in the cloud by Exa.

	🟩	**Device level configurations**: Several device-level configuration options are available, depending on the customers needs: Web proxies: ● The customer can opt to route their web traffic to SurfProtect by enforcing SurfProtect web proxies on devices. Deployment of proxy settings is managed by the customer, typically via Group Policy Objects for Windows, or MDMs for other devices. AD proxy: ● The use of the AD proxy requires additional configuration to be done on the customer's AD domain controller, which enables SurfProtect to associate user identities with web requests. PAC file: ● An alternative to direct proxy settings, customers can enforce SurfProtect captive portal by deploying a PAC file to all devices using their preferred method of deployment.
● Filtering Policy – the filtering provider publishes a rationale that details their approach to filtering with classification and categorisation as well as how the system addresses over blocking	🟩	Our filtering policy document is available here https://exa.net.uk/wp-content/uploads/2025/08/SurfProtect-Filtering-Policy-2025.pdf

- Group / Multi-site Management – the ability for deployment of central policy and central oversight or dashboard	🟩	As stated in the previous ‘Control’ question, MATs and Resellers have the ability to view and manage all their schools or customers via the SurfProtect panel.
- Identification - the filtering system should have the ability to identify users and devices to attribute access (particularly for mobile devices) and allow the application of appropriate configurations and restrictions for individual users. This would ensure safer and more personalised filtering experiences.	🟩	SurfProtect offers two different setups which enable user identities to be associated with web requests. **The AD proxy**: This option requires some configuration on the customer's AD domain controller, but is able to provide user identity data when Windows AD users make web requests. **Captive Porta**l: Available to school devices, BYOD devices and for devices taken out of school. The service presents users with a login prompt for the SurfProtect Captive Portal using AD credentials, a Google workspace account or a Microsoft Entra account. Using this service does require the associated user identity data to be uploaded to the Self Administration panel, either manually or using the automated integrations. **Configuring policies for specific users**: As described in the ‘Context appropriate differentiated filtering’ section, customers with access to the self administration panel are able to create custom filtering profiles. Each of these profiles can have a set of matching usernames or

	🟩	group names associated to them, which will then match any web request with matching user identity data
● 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 capability of their filtering system to manage content on mobile and web apps and any configuration or component requirements to achieve this	🟩	SurfProtect operates at the network level, intercepting and filtering all web traffic from browsers, applications and mobile apps, provided that the traffic is routed to SurfProtect. In the context of web-filtering, there is no fundamental differences between web traffic generated by traditional browsers, mobile browsers or mobile applications. As long as the site or application’s content is served via a web protocol (HTTP/HTTPS), it is treated the same by SurfProtect regardless of its source.
● Multiple language support – the ability for the system to manage relevant languages	🟩	SurfProtects search term filtering lists have access to a set of search terms curated by Exa. Each of these lists is also translated into the ten most used languages in the UK. SurfProtect is capable of categorising websites in languages other than English.
● 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	🟩	SurfProtect Quantum+'s PAC solution enables devices that are taken out of schools to be dynamically configured with a captive portal-enforcing proxy, enabling the filtering and association of web requests to users even when off the schools network.
● Reporting mechanism – the ability to report inappropriate content for access or blocking	🟩	There are a number of methods by which customers can report these issues, or update their filtering to allow

	🟩	or block access to content themselves: ● The Exa Networks support team are ready to help at any time with updating filtering settings. ● Both the Exa main website and our self-administration panels have a web chat functionality that allows for the submission of support requests and queries. ● Customers with access to the self-administration panels have access to manage their filtering settings, allowing them to allow or block any content they need to.
● Reports – the system offers clear granular historical information on the websites users have accessed or attempted to access	🟩	Customers with access to the self-administration panel are able to view SurfProtect analytics, which allows users to analyse all filtered web requests. SurfProtect also offers a real-time alerts feature that allows nominated contacts from the school to receive email alerts about specific user behaviour blocked by SurfProtect, such as visiting websites relating to/searching words related to sex, drugs or suicide.
● Safe Search – the ability to enforce 'safe search' when using search engines	🟩	By default SurfProtect enforces a 'Safe Search' setting. In the scenario the customers choose not to decrypt their HTTPS web

	🟩	traffic, SurfProtect DNS also redirects most well known search engines to their safe-search equivalents.
- Safeguarding case management integration – the ability to integrate with school safeguarding and wellbeing systems to better understand context of activity	🟩	Data from SurfProtect's historical and real-time analytics can be exported in CSV format and imported into these systems.

How does your filtering system manage access to Generative AI technologies (e.g. ChatGPT, image generators, writing assistants)?
In your response, please describe whether and how your system identifies, categorises, or blocks Generative AI tools; how access can be controlled based on age, risk, or educational need; any limitations in filtering AI-generated content—particularly where such content is embedded within other platforms or applications; and what support or configuration guidance you offer to schools to help them align with the UK Safer Internet Centre's Appropriate Filtering Definitions and relevant national safeguarding frameworks.

SurfProtect blocks access to generative AI technologies - such as ChatGPT - by default, through our dedicated 'Artificial Intelligence' category.

Filtering profiles within SurfProtect are highly customisable and can be applied at the level of individual users, user groups, IP ranges or proxy ports. This enables schools to provide differing access to this content based on age, educational need, or risk level. For example, pupils can have more restrictive profiles, blocking access to AI content entirely, while profiles for staff might allow access to approved tools.

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*".[1]

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

Beyond robust filtering, Exa, through The Exa Foundation, actively champions online safety education, empowering students, teachers and parents with the knowledge and skills needed to navigate the digital world safely and responsibly. Our initiatives compliment the technical safeguards we provide, fostering a culture of online safety from the ground up.

Examples of The Exa Foundation's Support for Wider Online Safety Education:

- **Engaging Pupil Workshops:** Our dedicated team visits schools across the UK to lead interactive and age-appropriate workshops. These hands-on sessions focus on equipping young people with practical strategies to protect themselves and others online. Key areas covered include:
    - Understanding digital footprints and online reputation.
    - Identifying and reporting fake news and misinformation.
    - Developing critical thinking skills to evaluate online content.
    - Knowing when and how to seek help if they encounter something worrying online.

[1] https://www.gov.uk/government/publications/keeping-children-safe-in-education--2

In the current academic year, The Exa Foundation has visited more than 100 schools in England, directly impacting thousands of pupils by fostering crucial digital literacy skills and empowering them to make safer choices online.

- **Staff Training and Resources:** Recognising the vital role educators play, The Exa Foundation also provides resources and, where requested, training for teachers to enhance their understanding of evolving online safety challenges and how to embed these lessons within the curriculum.

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	Katie Sinclair
Position	Governance, Risk and Compliance Officer
Date	12/09/2025
Signature	