

Appropriate Filtering for Education settings



June 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) 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	Talk Straight Ltd – T/A Schools Broadband
Address	Units 2-4, Backstone Business Park, Dansk Way, Ilkley, West Yorkshire, LS29 8JZ
Contact details	Telephone: 01133 222333 Email: info@schoolsbbroadband.co.uk
Filtering System	Netsweeper & FortiGuard Web Content Filtering
Date of assessment	October 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"> Are IWF members 	Netsweeper & FortiGuard	Netsweeper, Fortinet and Schools Broadband are long standing IWF members, supporting the IWF for over twelve (12) years.
<ul style="list-style-type: none"> and block access to illegal Child Abuse Images (by actively implementing the IWF URL list) 	Netsweeper & FortiGuard	The Netsweeper and Fortinet products integrate with the IWF CAIC illegal content list. The IWF functionality is not exposed in the graphical user interface and cannot be disabled.
<ul style="list-style-type: none"> Integrate the ‘the police assessed list of unlawful terrorist content, produced on behalf of the Home Office’ 	Netsweeper & FortiGuard	Netsweeper and FortiGuard integrate this list into their Web Filtering service and this is blocked by default.
<ul style="list-style-type: none"> Confirm that filters for illegal content cannot be disabled by the school 	Netsweeper	All categories associated with illegal content are locked at the system level and all school administrators cannot enable these categories.
	FortiGuard	Illegal content filters can only be disabled by an approved and designated administrator, as agreed by the senior leader team.

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.	Netsweeper	<p>Web Filter Category Hate Speech</p> <p>These sites portray views that are written, verbal, or illustrated, and are intentionally offensive to the public. The intent of these sites is to degrade, intimidate, or incite violent or prejudicial actions</p>

			<p>against individuals based on race, ethnic affiliation, nationality, gender, sexual orientation, religion, disability, or profession.</p> <p>Advocacy or instructional sites that promote the harming of individuals or groups and encourage or promote peer abuse, videos of physical assaults, written harassment and threats are also included.</p>
		<p>FortiGuard</p>	<p>Web Filter Category Discrimination</p> <p>Sites that promote the identification of racial groups, the denigration or subjection of groups, or the</p>
<p>Drugs / Substance abuse</p>	<p>displays or promotes the illegal use of drugs or substances</p>	<p>Netsweeper</p>	<p>Web Filter Category Substance Abuse</p> <p>These sites provide information about or promote the use of prohibited, illegal, controlled, or regulated substances for recreational rather than medicinal use. It can include sites that sell, encourage, or advocate the use of any substance that produces hallucinations, as well as the cultivation, manufacture, and distribution of any intoxicant and related paraphernalia.</p> <p>Informational sites that are clearly intended to provide descriptions of drugs and substances, their negative effects, and addiction potential are not included.</p>

		FortiGuard	<p>Web Filter Category Drug Abuse</p> <p>Websites that feature information on illegal drug activities including: drug promotion, preparation, cultivation, trafficking, distribution, solicitation, etc.</p>
Extremism	promotes terrorism and terrorist ideologies, violence or intolerance	Netsweeper	<p>Web Filter Categories Extreme, Hate Speech, Criminal Skills, Terrorism and Weapons</p> <p>Extreme - This includes sites that are considered far from normal and are categorised for their degree of intensity. The content features or promotes intentional, direct, and deliberate violence and destruction or the alteration of the human body and other living creatures. These sites may depict or promote torture, self-inflicted harm, mutilation, or other dangerous activities. Images and information that advocate and glorify eating disorders, suicide, death, gore, injuries, or sites that feature grotesque or frightening descriptions are also included.</p> <p>Hate Speech - These sites portray views that are written, verbal, or illustrated, and are intentionally offensive to the public. The intent of these sites is to degrade, intimidate, or incite violent or prejudicial actions against individuals based on race, ethnic affiliation, nationality, gender, sexual orientation, religion, disability, or profession.</p> <p>Advocacy or instructional sites that promote the harming of</p>

		<p>individuals or groups and encourage or promote peer abuse, videos of physical assaults, written harassment and threats are also included.</p> <p>Weapons - This includes sites that provide information related to the promotion, support, sale, or discussion of weapons and any related device used in combat that can injure or kill, such as guns, knives, or swords.</p> <p>Information on how to build weapons or bombs will also be included in 'Criminal Skills'.</p> <p>Terrorism - This includes content that promotes the illegal use, action or process of violence against civilians or a spectrum of society for the purpose of societal and/or political change.</p> <p>Criminal Skills - This includes sites with instructions or methods that promote, encourage, or provide skills considered to be illegal, criminal, violent or harmful to the public, and are forbidden by law. This can include questionable material and sites that promote nonviolent, unethical, or dishonest behaviour such as academic cheating, or software hacking/key breaking. This does not necessarily reflect the laws of any particular region or country.</p>
--	--	--

		FortiGuard	<p>Web Filter Category Extremist Groups</p> <p>Sites that feature radical militia groups or movements with aggressive anti- government convictions or beliefs.</p>
Gambling	Enables gambling	Netsweeper	<p>Web Filter Category Gambling</p> <p>This includes sites that encourage or provide information on the wagering or risking of money or any valuables on a game, contest, or other event in which the outcome is partially or completely dependent upon chance or on one's abilities. Sites that promote or facilitate gambling information, as well those that are purely factual and strategic sites that promote cheating are also included.</p> <p>This excludes sites that are clearly support sites for gambling addiction as well as travel destination sites that do not enable gambling.</p>
		FortiGuard	<p>Web Filter Category Gambling</p> <p>Sites that cater to gambling activities such as betting, lotteries, casinos, including gaming information, instruction, and statistics.</p>
Malware / Hacking	promotes the compromising of systems including anonymous browsing and other filter bypass tools as well as sites hosting malicious content	Netsweeper	<p>Web Filter Categories Malware, Hacking, Infected Hosts, Phishing, Viruses & Adware</p> <p>These categories block websites that are associated with hacking and malware.</p>

			<p>These are sites containing scripts, or code, that may be run in a hostile or intrusive manner to a system.</p>
		<p>FortiGuard</p>	<p>Web Filter Categories Malicious Websites & Hacking</p> <p>Malicious Websites - malicious content covertly downloaded to a user's machine to collect information and monitor user activity, and sites that are infected with destructive or malicious software, specifically designed to damage, disrupt, attack or manipulate computer systems without the user's consent, such as virus or trojan horse.</p> <p>Hacking - Websites that depict illicit activities surrounding the unauthorized modification or access to programs, computers, equipment and websites.</p>
<p>Pornography</p>	<p>displays sexual acts or explicit images</p>	<p>Netsweeper</p>	<p>Web Filter Categories Pornography & Nudity</p> <p>Pornography - This contains URLs that reference, discuss, or display pornographic images, videos, or other sexually oriented material that is created for the purpose of arousing sexual interest. Soft and hard-core pornography, sadomasochism, bestiality, fetishes, erotic stories, adult magazines, sex toys, or any other sexual related products are included.</p> <p>Nudity - This includes sites containing nude or semi-nude depictions of the human body. These depictions are not necessarily sexual in intent or effect, but may include sites</p>

			<p>containing nude paintings or photo galleries of an artistic nature.</p>
<p>Piracy and copyright theft</p>	<p>includes illegal provision of copyrighted material</p>	<p>FortiGuard</p>	<p>Web Filter Categories Pornography & Nudity and Risque</p> <p>Pornography - Mature content websites (18+ years and over) which present or display sexual acts with the intent to sexually arouse and excite.</p> <p>Nudity and Risque - Mature content websites (18+ years and over) that depict the human body in full or partial nudity without the intent to sexually arouse.</p> <p>Web Filter Categories Copyright Infringement, Peer to Peer and PIPCU.</p> <p>Copyright Infringement - This category contains sites that use, provide, or distribute information on copyrighted intellectual property or illicitly copied material which violates the owners' rights.</p> <p>Peer to Peer - This includes sites that distribute software and facilitate the direct exchange of files between users. Software that enables file searching, sharing and transferring across a network independent of a central server as well as web based sites of this nature are included.</p> <p>PIPCU - Sites in this category have been identified by the Policy Intellectual Property Crime Unit as containing potentially copyright infringing websites. This list is not</p>
		<p>Netsweeper</p>	

			<p>managed by Netsweeper. For more information, see the PIPCU website</p>
		FortiGuard	<p>Web Filter Category Peer to Peer</p> <p>File Sharing Websites that allow users to share files and data storage between each other.</p>
Self Harm	promotes or displays deliberate self harm (including suicide and eating disorders)	Netsweeper	<p>Web Filter Category Self Harm</p> <p>This category contains websites that depict and feature intentional and direct self-inflicted harm. Sites that advocate and glorify eating disorders, self-mutilation, and suicide are included in this category.</p>
		FortiGuard	<p>Web Filter Category Explicit Violence</p> <p>This category includes sites that depict offensive material on brutality, death, cruelty, acts of abuse, mutilation, etc.</p>
Violence	Displays or promotes the use of physical force intended to hurt or kill	Netsweeper	<p>Web Filter Category Extreme</p> <p>This includes sites that are considered far from normal and are categorised for their degree of intensity. The content features or promotes intentional, direct, and deliberate violence and destruction or the alteration of the human body and other living creatures. These sites may depict or promote torture, self-inflicted harm, mutilation, or other dangerous activities.</p> <p>Images and information that advocate suicide, death, gore, injuries, or sites that feature</p>

			grotesque or frightening descriptions are also included.
		FortiGuard	<p>Web Filter Category Explicit Violence</p> <p>This category includes sites that depict offensive material on brutality, death, cruelty, acts of abuse, mutilation, etc.</p>

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

Netsweeper

Netsweeper uses AI based technology for near real-time classification of dynamic website content over 90+ categories, including 47 different languages. This allows the platform to be flexible in customers selecting only the categories that need to be denied for each policy. Policies can be configured on a per user group basis, therefore the correct category restrictions can be enforced for the right user types.

Netsweeper also allows for local overrides of URL categorisation which can be applied to multiple policies. Therefore if a school needed a URL allowing for all users in a school, this is easily done with one (1) entry. Users with administrative rights are able to modify their policy (categories and lists) according to their permissions.

Schools Broadband applies a suite of policy templates of typical education sites and categories to ensure that schools do not experience over-blocking during implementation.

FortiGuard

General categorisation is based on an automated categorisation engine which has been developed in-house and which has evolved over more than 13 years since its initial conception.

The system uses language dictionaries to allow support in any language. Sites are scanned based on a number of methods:

- new pages on identified popular sites
- URLs which are requested by a user, but which are not rated. Such URLs will go into a queue to be rated based on hit count and the current charge on the system.
- Bulk requests from a specific customer. Such requests are treated case by case, but we generally offer this as a free service.
- Individual requests received from customers or users. These requests can be received in a number of ways (see below) and may be either requests to rate an unrated site, or requests to change the rating of a site.

In general, initial rating is done by the automated rating system. Malicious content (viruses, exploits) is not rated using this system (more details below) because such sites generally have legitimate visible content.

Ratings may also be obtained from third-party feeds, including feeds from governments or other organisations, containing such content as extremism or sexual violence.

Requests to change the rating of an already categorised URL will always be dealt with by a human, to ensure that the request gets the highest level of care and attention

Schools Broadband applies a suite of policy templates of typical education sites and categories to ensure that schools do not experience over-blocking during implementation.

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.

Schools Broadband retain log-file data for two years as per our GDPR policy for both Netsweeper and FortiGuard systems.

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

Dynamic content analysis as well as easy to manage unblocking and recategorisation tools ensure high levels of accuracy and ensure over-blocking is not a problem.

Schools Broadband also frequently reviews high-hitting blocked categories, URLs and keywords across the Schools Broadband estate to ensure educational content is allowed and over-blocking does not occur.

Filtering System Features

How does the filtering system meet the following principles:

Principle	Rating	Explanation
<ul style="list-style-type: none"> Context appropriate differentiated filtering, based on age, vulnerability and risk of harm – also includes the ability to vary filtering strength appropriate for staff 	Netsweeper & FortiGuard	<p>Users can be grouped in whatever way is required, and policies can be applied to different groups.</p> <p>By grouping users into differentiated groups, users will receive appropriate filtering to the group they are assigned.</p>
<ul style="list-style-type: none"> 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. 	Netsweeper & FortiGuard	<p>We combine both FortiGate and Netsweeper to help detect, identify and manage access to circumvention type technologies.</p> <p>This traffic can be identified alongside standard web traffic and the end user can choose to allow or block this type of circumvention traffic with standard firewall policies, layer-7 application</p>

		control, or domain/IP/hostname blocking per protocol type.
<ul style="list-style-type: none"> 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 	Netsweeper & FortiGuard	<p>Schools Broadband configure different administration roles to allow schools to manage and maintain the filtering and reporting themselves.</p> <p>All changes made by either the school or Schools Broadband are fully logged and auditable by all parties.</p> <p>Roles include, but not limited to:-</p> <ul style="list-style-type: none"> - Full Management - Policy Management - List Management - Reporting only
<ul style="list-style-type: none"> 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. 	Netsweeper	Netsweeper provides on the fly categorisation based on the content and context of text and links that appear on the page.
	FortiGuard	Content filtering can be enabled to block text based content present on a page. Content filters need to be populated with the text elements to be blocked based on regex or wildcard statements.
<ul style="list-style-type: none"> Filtering Policy – the filtering provider publishes a rationale that details their approach to filtering with classification and categorisation as well as over blocking 	Netsweeper	<p>The Netsweeper collective community together with AI technology and human oversight defines URL classification.</p> <p>Netsweeper publishes classification of filtering and categories on the Netsweeper website as well as a view in real time of new content categorised - https://www.netsweeper.com/live-stats/</p> <p>Netsweeper uses AI technology to categorise every URL. Netsweeper’s categorisation is real-time and combines a hierarchy of data (URL to category), with their Category Naming Service (CNS) as a global master database.</p> <p>If any customer anywhere in the world accesses a URL, that URL is</p>

		<p>submitted to the local policy server. If that policy server cannot find a category match, it is automatically submitted to the CNS and looked up there. If the CNS already has the category mapping it is immediately returned to the local system and cached there for future use. A policy decision is then made by the policy server.</p> <p>If neither the local system, nor the CNS has a category match, the URL is submitted to the Netsweeper "Artificial Intelligence" system that will interrogate the content at and around that URL, assess the content, detect if it references or contains malware, and assigns one or more categories to the URL into the CNS. This is then reported back to the local system, a policy decision is then made by the policy server.</p> <p>The CNS allows Netsweeper to adapt to trending URLs immediately due to its world-wide scope. If the local system hasn't seen a particular URL yet, then CNS probably has. If the URL has been assigned one or more categories, local systems see immediate responses.</p> <p>If the URL is truly "new" then the AI will typically process the content within a matter of seconds. The local policy servers can be configured with techniques to minimise the "new URL" wait.</p>
	<p>FortiGuard</p>	<p>Fortinet approaches web filtering differently for three broad areas: -</p> <p>Malicious content - This includes viruses and sites which are capable of exploiting vulnerabilities in users' browsers and other applications. Often these sites will be legitimate sites which have been compromised by a cybercriminal. The approach to detecting such sites is very different</p>

from general categorisation, since the visible content of the site provides no clues of the malicious content hidden within.

Offensive content - This includes such categories as pornography, violence and extremism, and are considered to be the categories which must be prioritised in terms of coverage and accuracy. As a result, a disproportionate amount of effort is given to rating these categories, in terms of human resources, research and development of automation tools, and ongoing daily processing.

General content - This includes such categories as shopping, news, sport etc, where ambiguities in rating can be tolerated. The goal of separating these groups is to ensure that the areas which represent the greatest risk are those for which Fortinet applies the highest priority. For the question of overblocking, care is taken to block on complete URLs wherever possible, rather than blocking based on a domain name or IP address. This approach allows a site to continue to function even if it contains malicious content, since only that content will be blocked, rather than the entire site being blocked because of one file. Note however that when a malicious file is identified on a given website, crawlers will be dispatched to try to identify any other malicious content which may be hidden in the same site.

However, sometimes it is appropriate to give a single categorisation to an entire domain, so a hierarchical search is used to allow entire subdomains or paths within a site to be blocked if necessary. This applies also to user defined URL patterns.

<ul style="list-style-type: none"> Group / Multi-site Management – the ability for deployment of central policy and central oversight or dashboard 	Netsweeper & FortiGuard	<p>We provide a single pane of glass service where policies can be shared across multiple schools.</p> <p>We also provide a central dashboard and interface for reporting against multi-site deployments and hierarchical views for Multi Academy Trusts.</p>
<ul style="list-style-type: none"> Identification - the filtering system should have the ability to identify users 	Netsweeper & FortiGuard	<p>Netsweeper and FortiGate integrates with existing directory systems such as Microsoft AD, LDAP, Apple LDAP, OpenLDAP, MS Azure AD and Google Workspace.</p> <p>For guest/wireless networks schools can utilise Radius to identify when a user has authenticated.</p> <p>Users can be passively identified by agents covering various platforms such as OSX, Google & Windows, or be identified through a captive portal for devices that are unable to have an agent installed.</p>
<ul style="list-style-type: none"> 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 	Netsweeper & FortiGuard	<p>Netsweeper and FortiGate are deployed as a dual network service. With this combination, Schools Broadband can detect and manage application protocols using DPI (Deep Packet Inspection).</p> <p>Standard web content delivered to applications can be filtered in the same manner as a web browser.</p> <p>For full Layer 7 application inspection, to facilitate full Mobile and app content inspection, Schools Broadband provide this service through the FortiGate platform.</p>
<ul style="list-style-type: none"> Multiple language support – the ability for the system to manage relevant languages 	Netsweeper	<p>Netsweeper provides real-time categorisation for 47 languages.</p>
	FortiGuard	<p>The Fortinet web filtering system has inherent multi-language support where each language has an extensive dictionary which is used by the rating system to categorise content. The human web filtering team has fluency in over 15 languages</p>

<ul style="list-style-type: none"> • Network level - filtering should be applied at 'network level' ie, 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) 	Netsweeper & FortiGuard	<p>Netsweeper and FortiGate filters are deployed as a network service.</p> <p>As the deployment is an inline network filter, we filter on all traffic transparently at the network level whilst a device is at school.</p>
<ul style="list-style-type: none"> • 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 	Netsweeper & FortiGuard	<p>Both solutions can facilitate the identification and filtering of all remote devices off the school network, using the same filtering policies that are provided in school.</p> <p>This comes in the form of a client application that can be installed on these remote devices. The software detects when working remotely and automatically starts filtering as if they the device was still inside the school without any user intervention.</p>
<ul style="list-style-type: none"> • Reporting mechanism – the ability to report inappropriate content for access or blocking 	Netsweeper & FortiGuard	<p>Designated contacts at each school have the ability to report inappropriate content via our customer portal, e-mail or via phone.</p> <p>Each school also has the ability to immediately block or allow access to any content if the user has correct management privileges.</p>
<ul style="list-style-type: none"> • Reports – the system offers clear historical information on the websites users have accessed or attempted to access 	Netsweeper & FortiGuard	<p>Schools Broadband has multiple reporting techniques outside of the vendor supplied reporting engines.</p> <p>Dashboards – These can be configured to show dynamically changing information for inappropriate content access requests or blocked requests. These can be customised per report admin so that they can see relevant information.</p> <p>Scheduled Reports – These are reports that can trigger e-mails on a set schedule.</p> <p>Instant Alerting – Using our dedicated Incident Management Platform, report admins can configure specific keyword or URL lists to get</p>

		instant notifications via Teams, E-Mail or Slack. Users can define their own lists, use our pre-defined lists (created in conjunction with 3 rd party agencies), or use a hybrid of the two.
<ul style="list-style-type: none"> • Safe Search – the ability to enforce ‘safe search’ when using search engines 	Netsweeper & FortiGuard	Safe Search can be enforced on a per user group basis.

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:


Schools Broadband can provide custom messaging to students when they attempt to access specific denied content. For example: Schools Broadband can provide an information page on hate speech, drug abuse, viruses, etc rather than just simply denying them. This allows schools to message specific topics and provide students with more information on how and where to obtain help.

Schools Broadband is also looking to provide on-line self-paced training courses that could be utilised by school users. This would include not only safeguarding, but also cyber security in general to help support the broader obligations within schools.

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	Luke Watson
Position	Product Manager
Date	19-October-2023
Signature	

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