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	iboss
Address	5-11 Worship St, London EC2A 2BH
Contact details	+44 20 3884 0360 emeia@iboss.com
Filtering System	iboss
Date of assessment	02/10/2025

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		Member Since 2013
• and block access to illegal Child Abuse Images (by actively implementing the IWF URL list), including frequency of URL list update		Yes – CAIC list is in a restricted Category
• Integrate the 'the police assessed list of unlawful terrorist content, produced on behalf of the Home Office'		The List is integrated and can be locked.
• Confirm that filters for illegal content cannot be disabled by anyone at the school (including any system administrator).		This can be configured within the platform to only allow a break glass account the ability to disable these controls

Describing how, their system manages the following illegal content

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.		Placed into the Pornography (Child) category
controlling or coercive behaviour	Online actions that involve psychological abuse, manipulation, or intimidation to control another individual, often occurring in domestic contexts.		Placed into violence and hate category / Extreme category
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.		Placed into the violence and hate / pornography category
extreme pornography	Pornographic material portraying acts that threaten a person's life or could result in serious injury, and is deemed obscene and unlawful.		Placed into violence and hate category / Extreme category
fraud	Deceptive practices conducted online with the intent to secure unfair or unlawful financial gain, including phishing and scam activities.		Placed into the scams / phishing and suspicious category
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.		Placed into Violence and Hate category

inciting violence	Online material that encourages or glorifies acts of violence, posing significant risks to public safety and order.		Placed into Violence and Hate category
illegal immigration and people smuggling	Content that promotes or facilitates unauthorized entry into a country, including services offering illegal transportation or documentation.		Placed into Violence and Hate category
promoting or facilitating suicide	Material that encourages or assists individuals in committing suicide, posing serious risks to vulnerable populations.		Placed into suicide category
intimate image abuse	The non-consensual sharing of private sexual images or videos, commonly known as "revenge porn," intended to cause distress or harm.		Placed into the pornography / Violence and Hate category
selling illegal drugs or weapons	Online activities involving the advertisement or sale of prohibited substances or firearms, contravening legal regulations.		Place into the controlled drugs / weapons category
sexual exploitation	Content that involves taking advantage of individuals sexually for personal gain or profit, including trafficking and forced prostitution.		Placed into the pornography / Violence and Hate category
Terrorism	Material that promotes, incites, or instructs on terrorist activities, aiming to radicalise individuals or coordinate acts of terror.		Placed into the terrorism & radicalisation category

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		Placed into the gambling category
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		Placed into the Violence and Hate category
Harmful content	Content that is bullying, abusive or hateful. Content which depicts or		Placed into the Violence and Hate category

	encourages serious violence or injury. Content which encourages dangerous stunts and challenges; including the ingestion, inhalation or exposure to harmful substances.		
Malware / Hacking	promotes the compromising of systems including anonymous browsing and other filter bypass tools as well as sites hosting malicious content		Placed into the malware category / captured by malware defence engines.
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		Placed into the Suspicious category
Piracy and copyright theft	includes illegal provision of copyrighted material		Placed into the Piracy Category
Pornography	displays sexual acts or explicit images		Placed into the pornography category
Self Harm and eating disorders	content that encourages, promotes, or provides instructions for self harm, eating disorders or suicide		Placed into the self-harm category
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.		Placed into the Violence and Hate category

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

Application (Layer 7), controls such as Games, Chat, IM, P2P, command line tools, etc. - Layered Google service controls (Safe Search, Safe image Search, YouTube and Gmail controls) - Deep Packet Inspection (DPI), for evasive applications such as Tor, BitTorrent, Ultra surf, Psiphon etc. - Browser and OS controls - File extension and file type download controls - Social Media Controls (Facebook, Twitter, Pinterest etc) - Port Blocking - Sleep Schedules - Keywords with high-risk real-time alerting - Real-time monitoring - CASB App discovery - AI based cloud app discovery, categorisation and risk analysis - YouTube Libraries - AI service control, monitoring including academic dishonesty detection)

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.

Iboss has multiple levels of log retention to suit the requirements of the organisation. These are configured in storage sizes to ensure logs can be retained to meet the requirement. The iboss logs each connection made by a user and has powerful reporting tools to provide such reports as User Risk, Search term alerting and classification, real time views and drill down and email reports

iboss also includes education specific reporting including User Behavior and Risk – Shows a summary of high-risk user behaviors and interactions, including high risk, suicide, self harm , threats to harm, productivity loss, liability, etc. Admins can drill into surfaced user behaviour related to domain visits, domains blocked, keywords searched, etc.

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

All categories have 4 modes, Allow, Block, Stealth (Stealth mode can be used for content monitoring without blocking content) and Soft-Block (Warning the user a website is within a category and allowing them to click through the block page and reach the site while recording that action in logs) All categories also have priorities so that categories can be weighted appropriately for the policy type or age rating. For example, if the games category is priority 0 and blocked, and the education category is priority 1 and allowed – game web sites with no education content will only be placed into the games category and therefore blocked. However, game web sites with educational game content will be placed into both the games and education category, and as the education category has a higher priority and is allowed, the educational game web sites will be allowed. content without intervention from the web filter administrator. Block pages can have ‘exceptions per policy’. This allows for feedback to be sent to the filtering administrator, directly from the block page, including a reason why the web site should be unblocked. Exceptions can generate real-time alerts and have their own administration area for easy unblock/block tasks. Uncategorized URL’s can be blocked, blocked with override controls and are per policy

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		Policies can be per year, group, or classroom with weighted categories.
• 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.		Full analysis of all TCP and UDP ports with algorithms to detect, block and quarantine endpoints running evasive applications such as: - Tor - Psiphon - Ultrasurf - OpenVP - BitTorrents - Chat Apps - + Others
• 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		All controls are intuitive with context sensitive online help and can provide delegated access to teaching staff. All controls are via a reactive web console that fits to any screen size
• 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.		Iboss is able to apply contextual filtering via scanning for triggers and alerts across web access from the end user. AI generated content would also be scanned, LLM content can also be recorded for further analysis by safeguarding teams. This includes inline warnings as well as conversation recordings. All iboss filtering is performed in realtime.
• 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		There are many ways to deploy the iboss platform within a customer environment. These include: • Agent based for managed endpoints • DNS filtering for low touch minimal risk guest networks • BYOD devices can connect using Guest Sessions/Remote Browser Isolation (preferred) or

		Reverse Proxy. While requests going through a reverse proxy can provide all the security functions, it does not prevent data from landing on non-enterprise-owned assets. • Explicit Proxy (client-aware, onsite deployments) - Explicit configuration to direct request flow via proxy-aware applications. • Transparent Mode (non-client aware, onsite deployments) - Configuration does not have to be performed on the individual endpoints. Policy compliance is achieved by the PEPs inspecting traffic flows egressing through their natural path. Transparent Mode secures non-organization-owned assets or non-conforming devices that cannot accept a cloud connector or proxy policy. All options are support both for hardware and full cloud deployments
● 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		As a global company iboss provides filtering and malware defence solutions with highly configurable controls so as to meet the various governance and compliance regulations in different countries.

● Group / Multi-site Management – the ability for deployment of central policy and central oversight or dashboard		The iboss cloud management platform provides a 'Single Pane of Glass' management view. The node based architecture of the iboss Distributed Gateway Platform allows filtering gateways to be located anywhere and then cloud joined for centralized management and reporting. The role based and delegated administration model of the platform means that multiple administrators can manage the gateways, and ISP's or MSP's can manage multiple environments and accounts from a single console.
● 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.		The iboss SWG integrates with multiple directory and SSO environments including but limited to: Active Directory , SAML, Radius (802.1x, Wireless, NAC), Edirectory, OpenDirectory, LDAP, Cloud Identity such as google, Okta & Azure AD, and has options for BYOD and 'nondomain' joined devices (iOS)
● 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		The iboss SWG inspects all web streams (all TCP and UDP ports), and has full visibility of bidirectional web traffic from any type of web application not just web browsers. This allows the SWG to have granular controls for mobile, guest and BYOD devices including non-browser based applications.
● Multiple language support – the ability for the system to manage relevant languages		Content can be categorized in any language and logging and keyword controls accept any character set (Unicode).
● Remote devices – with many children and staff working remotely, the ability for school		The iboss cloud is able to protect devices from any location with the use of the

owned devices to receive the same or equivalent filtering to that provided in school		iboss cloud connector. This cloud connector is able to redirect traffic to the platform regardless of location and apply the same level of protection and filtering
● Reporting mechanism – the ability to report inappropriate content for access or blocking		The iboss SWG has an inbuilt micro SIEM known as the 'Reporting and Analytics console'. This separated reporting console has realtime reporting , and monitoring, query reports , drill down reports, and scheduled reports. In addition, realtime alerting can be triggered on keywords, attempted access to blocked categories, or use of evasive or high risk applications (plus device quarantine) and User Risk Reporting
● Reports – the system offers clear granular historical information on the websites users have accessed or attempted to access		URL and Event logging is via the iboss 'Reporting and Analytics' console that includes granular historical reporting that is customizable and exportable into popular formats (HTML, CSV, PDF etc). Reporting to external systems such as SIEM's is also supported via API or Syslog.
● Safe Search – the ability to enforce 'safe search' when using search engines		Iboss is able to enforce 'safesearch' or equivalent tools on major search engines.
● Safeguarding case management integration – the ability to integrate with school safeguarding and wellbeing systems to better understand context of activity		Iboss operates both email altering for real time integration as well as the open API allowing for full integration with 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.

Iboss enables the safe use of LLM and Generative AI tools. This can offer the ability to coach and warn users about AI risk, monitor and record conversation with LLM tools, these tools can create events and triggers to provide safeguarding alerts. These controls are policy driven and can be targeted to make them age appropriate. These are also bundled onto our existing tools and controls to offer safesearch enforcement. HTTPS content inspection, real time filtering and control, student risk module and data leak analysis and reporting.

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

iboss produces weekly blogs and other media regarding online safety and current threat vectors, along with advice to keep networks and their users safe

[1] 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	Anthony Robinson
Position	Sales Engineering Team Lead, EMEIA
Date	02/10/2025
Signature	*A. Robinson*