Appropriate Monitoring for Schools

May 2025

Monitoring 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 monitoring providers to illustrate to schools how their particular technology system(s) meets the national defined ‘appropriate monitoring’ standards. Fully completed forms will be hosted on the UK Safer Internet Centre website alongside the definitions

The results will help schools better assess, in conjunction with their completed risk assessment, if the monitoring system is ‘appropriate’ for them.

Company / Organisation	TrustLayer
Address	Belvedere House 4.2, Basing View, Basingstoke, RG21 4HG
Contact details	support@trustlayer.co.uk 08452309590
Monitoring System	TrustLayer CloudUSS (Web & Cloud Security)
Date of assessment	5 September 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.	🟨

Monitoring Content

Monitoring providers should ensure that they:

Aspect	Rating	Explanation
● Are IWF members	🟧	Indirectly, our threat intel partner for URL classification (zvelo) is a long-term member.
● Utilisation of IWF URL list for the attempted access of known child abuse images	🟩	
● Work with CTIRU 'the police assessed list of unlawful terrorist content, produced on behalf of the Home Office'	🟩	
● Confirm that monitoring for illegal content cannot be disabled by anyone (including any system administrator) at the school	🟩	

Illegal Online Content

Monitoring providers should both confirm, and describe how, their system monitors/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.	🟩	The domain and URL monitoring service contains over 500 web categories and granular sub-categories. Top level categories are listed here: https://help.clouduss.com/product-web-security/web-categories-list Custom lists, keywords, and RegEx filters can also be defined. ML algorithms analyse 'unclassified' or new URLs in real-time (default blocked).
controlling or coercive behaviour	Online actions that involve psychological abuse, manipulation, or intimidation to control another individual, often occurring in domestic contexts.	🟩	
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.	🟩	
extreme pornography	Pornographic material portraying acts that threaten a person's life or could result in serious injury, and is deemed obscene and unlawful.	🟩	
fraud	Deceptive practices conducted online with the intent to secure unfair or unlawful financial gain, including phishing and scam activities.	🟩	
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.	🟩	

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

Inappropriate Online Content

Monitoring providers should both confirm, and describe how, their system monitors/manages the following content

Content	Explanatory notes – Content that:	Rating	Explanation
Gambling	Enables gambling	🟩	The domain and URL monitoring service contains over 500 web categories and granular sub-categories. Top level categories are listed here: https://help.clouduss.com/product-web-security/web-categories-list Custom lists, keywords, and RegEx filters can also be defined.
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.	🟩	
Hate speech / Discrimination	Content that expresses hate or encourages violence towards a person or group based on something such as race, religion,	🟩	

	sex, or sexual orientation. . Promotes the unjust or prejudicial treatment of people with protected characteristics listed in the Equality Act 2010	🟩	ML algorithms analyse ‘unclassified’ or new URLs in real-time (default blocked).
Malware / Hacking	promotes the compromising of systems including anonymous browsing and other filter bypass tools as well as sites hosting malicious content, including ransomware and viruses	🟩	
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	🟩	
Pornography	displays sexual acts or explicit images	🟩	
Self Harm and eating disorders	encourages, promotes, or provides instructions for self harm or eating disorders	🟩	
VAWG	Promotes or glorifies violence, abuse, coercion, or harmful stereotypes targeting women and girls, including content that normalises gender-based violence or perpetuates misogyny.	🟩	

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

Highly granular rules and policies can be curated to create powerful filtering on domain, URL, category etc. Unclassified content is analysed autonomously and web classifications updated in real-time.
The Cloud Application module extends this to full http/https interception and analysis to track activity inside cloud-based SaaS applications, including AI tools.

Monitoring System Features

How does the monitoring system meet the following principles:

Principle	Rating	Explanation
• Age appropriate – includes the ability to implement variable monitoring appropriate to age and vulnerability. This will in turn define which alerts are prioritised and responded to. Further situations may warrant additional capability, for examples boarding schools or community based access	🟩	Enforcement of Safe Search functions etc. Policies can be defined based on device attributes, directory org units, browser, time, geo-location etc.

- Alert Management – how alerts are managed – if schools manage system alerts or support/management is provided	🟩	Real-time log streaming to 3rd party SIEM/SOAR.
- Audit – Any changes to the monitoring system are logged enabling an audit trail that ensures transparency and that individuals are not able to make unilateral changes.	🟩	Immutable audit trail for all logins and actions carried out
- BYOD (Bring Your Own Device) – if adopted by the school and the system includes the capability to monitor personal mobile and app technologies (ie not owned by the school), how is this deployed and supported and how data is managed. Does it monitor beyond the school hours and location	🟩	If connectivity enforced via on-prem gateway any device can be enforced via custom policies. Routing via captive portal or transparent proxy. 24/7 filtering and reporting.
- Data retention –what data is stored, where is it (physically– ie cloud/school infrastructure) stored and for how long. This should also include any data backup provision	🟩	Client defines storage region and data sovereignty. Logs stored on regional datacentre (AWS) and separated by tenant. Options for data redaction. Encrypted in flight (TLS ≥1.2) and at rest (AES56). Logs retained for 30 days for query in portal, real-time streaming available and scheduled exports.
- Devices – if software is required to be installed on devices, the monitoring system should be clear about the devices (and operating systems) it covers	🟩	Agent deployment to Windows/Mac/Chromebook, or on-prem gateway if device unmanaged. Policies can be tailored per OS, Browser Agent etc.
- Flexibility - changes in keywords (addition or subtraction) can be easily amended according to an agreed policy	🟩	Keywords and RegEx controls easily configured and updated in web-based portal.
- Group / Multi-site Management – the ability for deployment of central policy and central oversight or dashboard	🟩	Fully multi-tenant, Entra/AzureAD integration, Google Cloud etc. with templates and global policies.
- Harmful Image detection – The inclusion or extent to which visual content is discovered, monitored and analysed (e.g. Image hash).	🟨	Feature deprecated due to false positives. Still use QR code analysis, fake login pages etc.
- Identification - the monitoring system should identify users and devices to attribute activity (particularly for mobile devices) and ensure the application of appropriate configurations for individual users.	🟩	Agent supports full device and user identity with additional attributes available.

• Monitoring Policy – How are all users made aware that their online access is being monitored? Is any advice or guidance provided to support schools?	🟩	System tray icon and toast messages, custom branded templates to Warn/Block browser activity. Dislcaimer templates. Detailed KB articles and Customer Success to help with policy fine-tuning.
• 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 monitoring system operate across mobile devices and app content. Providers should be clear about the capability of their monitoring system to monitor content on mobile and web apps and any configuration or component requirements to achieve this.	🟩	Onsite traiifc can be routed through local gateway with captive portal or transparent proxy to secure mobile devices with no agents. Full cloud-hosted mobile private relay available Jan 2026 for Android and IOS via native VPN functionality.
• Multiple language support – the ability for the system to manage relevant languages?	🟩	Global databases for content matching.
• Prioritisation – How are alerts generated and prioritised to enable a rapid response to immediate issues. What operational procedures are in place to facilitate that process?	🟩	Log data streamed real-time to SIEM/SOAR system for alerting and response.
• Remote monitoring – with many children and staff working remotely, the ability, extent and management for the monitoring of devices (school and/or personal). Included here is the hours of operation together with the explicit awareness of users. Monitoring should focus on school-owned and managed devices. When shared devices are used, schools must ensure users log in individually. This allows monitoring systems to apply restrictions and configurations based on user profiles, improving the safeguarding process.	🟩	Remote devices managed via deployed agent and receive full protection as if on site. Policies can be controlled on a time-quota (i.e. relaxing/tightening social media access between certain hours or days)
• Reporting – how alerts are recorded within the system?	🟩	Full and detailed Analytics and Reporting engine for trend analysis, incident response, and forensic investigation. Every single http and https request and response is logged with granular metadata.
• Safeguarding case management integration – the ability to integrate with school safeguarding and wellbeing systems to better understand context of activity	🟩	Logs can be fed to 3rd party SIEM. Reporting can be configured to give trends on behaviour.

Pro Active Monitoring - how any pro-active monitoring support is provided including if any automation is utilised and the safeguarding capability of the organisation's teams.

Please note below opportunities or enhancements to support schools (and other settings) with their obligations around Keeping Children Safe in Education?

TrustLayer also offers a Security Awareness Training platform to educate users on safe internet usage, appropriate behaviour and digital safety (i.e. social media and social engineering)

How does your monitoring system identify and respond to activity involving Generative AI technologies (e.g. AI prompts, content creation, or platform use)?

In your response, please explain how your system captures or analyses user interactions with Generative AI tools; to what extent logs or alerts reflect potential safeguarding risks associated with AI-generated content (such as harmful prompts or inappropriate use of image and text generation); any known limitations—whether technical, privacy-related, or device-specific—that may affect your system's ability to monitor such activity; and what guidance you provide to schools to support their understanding and management of Generative AI-related risks.

Any interaction with a SaaS application via a web browser can be controlled in a granular fashion. We have a catalogue containing thousands of applications and dozens of AI tools. Apps are fingerprinted and classified via our automated profiling engines and manual verification – this means individual actions within Gen-AI tools can be controlled (such as limiting file uploads or specific prompt types). Each action within a SaaS application is allocated a baseline 'Risk' attribute which is displayed to the system admin and can be overridden based on individual preferences.

Monitoring 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	Gareth Lockwood
Position	CTO
Date	5/9/25
Signature	