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	Utropolis
Address	Mansfield, Dalkeith, EH22 5TJ
Contact details	sales@utropolis.io
Filtering System	Utropolis
Date of assessment	02/02/2025

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		Utropolis are proud to be members of the IWF and we support their work.
• and block access to illegal Child Abuse Images (by actively implementing the IWF URL list), including frequency of URL list update		The IWF list is an integral part of the protection we provide to our users.
• Integrate the 'the police assessed list of unlawful terrorist content, produced on behalf of the Home Office'		We actively block access to all sites on the CTIRU block list. Updates and changes to this list are integrated into Utropolis as soon as they become available.
• Confirm that filters for illegal content cannot be disabled by anyone at the school (including any system administrator).		The IWF and CTIRU block lists are enabled by default and cannot be disabled by anyone at the school.

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.		In addition to the IWF block list, Utropolis utilises selected open-source lists, URL keywords, search term and context-specific text analysis to block this kind of content.
controlling or coercive behaviour	Online actions that involve psychological abuse, manipulation, or intimidation to control another individual, often occurring in domestic contexts.		Contextual analysis of text patterns identifies, blocks and alerts safeguarding teams when such behaviour is identified.
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.		A combination of careful research, filtering technology and open-source block lists prevents content of this kind from being accessible.
extreme pornography	Pornographic material portraying acts that threaten a person's life or could result in serious injury, and is deemed obscene and unlawful.		All pornography and sexual material is blocked by default on the Utropolis platform.
fraud	Deceptive practices conducted online with the intent to secure unfair or unlawful financial gain, including phishing and scam activities.		We maintain categories and URL lists to prevent access to material such as fraudulent 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.		Utropolis has several categories enabled by default to identify and prevent these kinds of hate speech from being accessible.
inciting violence	Online material that encourages or glorifies acts of violence, posing significant risks to public safety and order.		Content which incites violence is identified through our contextual text analysis tools and classified into appropriate categories so that safeguarding teams can challenge these attitudes effectively.
illegal immigration and people smuggling	Content that promotes or facilitates unauthorized entry into a country, including services offering illegal transportation or documentation.		Using our extensive and evolving URL block lists and text analysis tools, Utropolis prevents users from engaging with this type of material online.
promoting or facilitating suicide	Material that encourages or assists individuals in committing suicide, posing serious risks to vulnerable populations.		Utropolis delivers protection from users seeing this kind of content and provides immediate safeguarding notifications.
intimate image abuse	The non-consensual sharing of private sexual images or videos, commonly known as "revenge porn," intended to cause distress or harm.		Through our network of block lists and contextual analysis, Utropolis actively monitors and prevents access to content such as this.
selling illegal drugs or weapons	Online activities involving the advertisement or sale of prohibited substances or firearms, contravening legal regulations.		Preventing access to illegal items is an integral part of the Utropolis offering and is managed through keyword and URL analysis as well as block lists.
sexual exploitation	Content that involves taking advantage of individuals sexually for personal gain or profit, including trafficking and forced prostitution.		Utropolis monitors for this kind of content, blocks it and reports attempted access to safeguarding teams.
Terrorism	Material that promotes, incites, or instructs on terrorist activities, aiming to radicalise individuals or coordinate acts of terror.		In addition to the implementation of the Counter-Terrorist Internet Referral Unit blocklist, Utropolis uses text analysis and reporting mechanisms to keep users safe and help keep schools in compliance with Prevent.

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		Utropolis technology prevents Gambling and Gambling-related sites from being accessible through a specific category.
Hate speech / Discrimination	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		The Utropolis filter is designed to identify and block a wide range of hate speech types before providing accurately categorised and timely information to safeguarding teams. It is designed to discriminate between hate speech and content which combats that kind of discrimination so that it can allow supportive content to be accessed.
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.		Through our continual research and use of open-source intelligence as well as information on historical trends, the Utropolis filter is kept up to date with the ability to block emerging trends.
Malware / Hacking	promotes the compromising of systems including anonymous browsing and other filter bypass tools as well as sites hosting malicious content		There are strong protections in place on Utropolis to prevent circumvention of our filter. This ties into customer device management platforms.
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		Open-source intelligence is utilised to prevent access to sites which frequently espouse mis-/dis-information. We also employ active text analysis to detect frequently used keywords and text patterns which indicate this kind of content.
Piracy and copyright theft	includes illegal provision of copyrighted material		This type of content is effectively managed through a combination

			of our filtering processes and open-source block lists.
Pornography	displays sexual acts or explicit images		Pornographic and sexual content is proactively blocked through the use of URL bock lists and also identified on the fly through careful contextual analysis filtering. All such activity is immediately notified to designated school safeguarding personnel.
Self Harm and eating disorders	content that encourages, promotes, or provides instructions for self harm, eating disorders or suicide		Drawing upon our research and algorithms, the Utropolis filter effectively blocks this kind of content and reports attempted access to the safeguarding team. Our system is able to identify sites aimed at supporting users seeking information about this topic from those encouraging or glorifying it.
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.		Utropolis has an extensive category dedicated exclusively to identifying Gender-Based Violence in all its forms as well as a counter-category to identify content providing support for people experiencing this kind of abuse.

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

Through a unique and highly-tuned combination of URL-based, keyword-based, search term driven and contextual text analysis, Utropolis monitors and identifies content across over 140 categories. To develop and maintain our capabilities, we draw upon our extensive knowledge as educators and veterans of the web-filtering and EdTech industry to carry out relevant up-to-date research, in addition to drawing upon carefully selected open-source intelligence and URL block lists. We also utilise the capabilities offered by counter-terrorist policing and the Internet Watch Foundation to ensure our users and our filters are compliant with modern legislation and standards such as KCSIE.

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.

Utropolis maintains 90 days of user logfiles in an encrypted format which is easy for school administrators to access through our dashboard. We then store this for 1 year in an encrypted offline format to aid any investigations that the school may require before safely disposing of it.

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

Many of our categories are matched by counter-categories which provide help and support for people facing those specific challenges. Our system is carefully tuned and tested to balance protection with usefulness, and we have internal systems to check the impact of any changes before they are deployed. We also have a highly reactive system including the integration of a proprietary AI tool to correct any instances of over-blocking that may occur.

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		Our system is designed to be strong enough that it can be relied on to block illegal content, whilst remaining flexible enough that it can be easily adapted by school administrators to suit the changing needs of mainstream and SEND pupils from nursery right through college and university ages and provide separate levels of protection for staff if required.
• 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.		We have tested Utropolis against a range of common circumvention techniques and we are confident in the system's ability to remain in control of web traffic. We actively block all mainstream services which could facilitate filter circumvention and proactively monitor for attempts to avoid the system.
• 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		We have an innovative interface to make this simple and intuitive for schools.
• 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'		All traffic is deeply analysed to ensure that hidden content cannot make it past the filter. Utropolis monitors all incoming text and code and actively analyses it for signs of illegal and undesirable content as it

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.		travels to the user device, blocking it before it even arrives if necessary.
• 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		Utropolis is configured remotely on managed devices through schools existing device management systems. This means there is zero impact on school devices and systems and secure protection is implemented no matter how the device is connected to the internet.
• 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 categories and rationales are available to our clients alongside expert assistance in deciding the best solutions for the needs of the institution. A reporting button appears on every block page for staff to request unblocking of a website. User documentation is available within the product to inform decisions over category access.
• Group / Multi-site Management – the ability for deployment of central policy and central oversight or dashboard		Our in-development multi-site dashboards and management tools will make oversight of usage and deployment of filtering policies easy for administrators.
• 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.		Due to our integration with existing device management tools, schools are able to deploy policies and track usage and alerts down to the individual user level.
• Mobile and App content – mobile and app content is often delivered in entirely different mechanisms from that delivered through a		When combined with responsible device management tool usage,

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		Utropolis provides a robust level of protection from harm online. All app data is processed through our filter.
• Multiple language support – the ability for the system to manage relevant languages		We are currently developing a robust framework to support multiple languages. While our filtering is presently optimised for English, multi-language integration is a priority roadmap objective to ensure inclusive and comprehensive protection across diverse educational environments.
• 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		The Utropolis filter configuration is deployed to the device so it operates with full functionality regardless of where the device is or what network it is connected to.
• Reporting mechanism – the ability to report inappropriate content for access or blocking		There is a clear button provided on the block screen to report inaccurately blocked content. Utropolis can also be contacted through school administrators to request access or block content.
• Reports – the system offers clear granular historical information on the websites users have accessed or attempted to access		There is a detailed graphical dashboard allowing administrators to gain insights into individual usage and events or overall usage trends.
• Safe Search – the ability to enforce 'safe search' when using search engines		The integration of Safe Search is an important component of the Utropolis filter.

• Safeguarding case management integration – the ability to integrate with school safeguarding and wellbeing systems to better understand context of activity		Safeguarding alerts are highly configurable, can be delivered in a number of ways and can be sent to as many individuals as required. This means granular, relevant alerts are easily set up and the people who need the information can get it quickly.

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.

Utropolis dynamically identifies and categorises Generative AI technologies into a dedicated "AI Tools" classification. This allows schools to implement granular access controls tailored to their specific safeguarding requirements and educational needs. In alignment with UKSIC's Appropriate Filtering principles, administrators can apply group-based policies, allowing access for those whom it will benefit while restricting it for users where it is judged inappropriate.

Where AI functionality is natively integrated into an encrypted application or productivity suite, we advise schools to manage these features via their respective Mobile Device Management (MDM) or software administrative portals. Utropolis remains highly responsive to these challenges, providing bespoke support and testing to help schools make informed, evidence-based judgement calls.

We provide comprehensive configuration guidance to help schools align their AI policies with DfE and UK Safer Internet Centre frameworks. By maintaining a flexible, category-based approach, we empower DSLs to balance the risks associated with Generative AI, such as data privacy and content reliability, against its evolving role in the modern curriculum.

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

At Utropolis, we recognise that technical filtering is most effective when it serves as a complement to, rather than a replacement for, high-quality digital literacy education. We view the

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

relationship between the provider and the school as a collaborative partnership; while our system dynamically prevents access to harmful or distracting content, we acknowledge that no filter is infallible in an ever-evolving digital landscape. Our goal is to work as part of the school's "safeguarding team," responding dynamically to emerging trends to maintain a safe learning environment.

We believe online safety is a fundamental life skill—comparable to a child learning to navigate physical spaces independently. Consequently, we advocate for a curriculum-led approach to safety and recommend resources such as Common Sense Education as an excellent foundation for building digital citizenship.

To directly support this, Utropolis is currently developing a suite of age-appropriate introductory modules for students. These lessons explain the role of filtering in a school context and empower users to take an active role in their own safety. By teaching students how to respond to inappropriate content in transferable, age-appropriate terms, we help them develop resilience that extends beyond the school network. Backed by a team with decades of teaching and safeguarding experience, we provide schools with robust reporting systems and expert guidance to ensure our technology supports a truly broad and balanced 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	Jonathan Henderson
Position	Cybersecurity and Content Analyst
Date	01/05/2026
Signature	J D Henderson