



Smoothwall Guidance on Generative AI for Schools

Generative AI is changing the nature of online content. Static, pre-curated web pages are being replaced by dynamic interactions in the form of conversations, and real-time generated customised content. The AI tools driving this shift are also constantly innovating and altering the way they work.

In this rapidly changing environment, web filtering alone cannot guarantee that students will be protected from harmful content produced by AI tools. We therefore strongly recommend that education settings combine real-time, content-aware filtering with robust digital monitoring to maximise student safety.

It is also imperative for education settings to perform appropriate risk assessments of any AI tool (particularly those presented in a chatbot type interface) they intend for students to access. These assessments should consider potential safeguarding risks, as well as any data privacy and intellectual property implications. Further guidance on this process is available in the DfE policy paper *Generative AI in Education*.



How Smoothwall supports the safe, responsible use of AI tools

Unmatched real-time filtering

Generative AI tools produce content dynamically, in real time. The only form of filtering able to block harmful material produced in this way is that which performs real-time, dynamic content analysis. Smoothwall Filter remains the only 100% real-time filter on the market.



Granular control of access to AI tools

Smoothwall Filter's "Unsafe AI Chat" category ensures that certain AI tools known to be dangerous are automatically blocked. With the broader "AI Tools" category, policies can restrict access to all AI tools except for those explicitly permitted for use. Tailored policies also offer the flexibility to permit or deny access to specific AI tools based on factors such as year group.

UK SIC-accredited protection

Smoothwall Filter is accredited by the UK Safer Internet Centre, whose latest guidelines state that schools should consider the extent to which their filter can dynamically analyse AI-generated content in real time. DNS, URL and non-real-time filters are unable to do this at any level. Smoothwall Filter meets the standard that **"content is dynamically analysed as it is streamed in real time to the user."**

However, given the changing nature of how AI content is produced and delivered to users, even advanced filters will experience limitations in this area. This is why best level filtering needs to be combined with risk assessments and effective monitoring to maintain appropriate protection for students.

Essential digital monitoring

AI Tools are interactive, which means digital monitoring is critical to protect students from the associated safeguarding risks. Responses produced by these tools are shaped by user input - only digital monitoring can fully assess these entries. Smoothwall Monitor's keystroke analysis can detect potential risks in what students type into generative AI tools in real time and alert schools accordingly.

The extra layer of protection provided by our trained human moderators dramatically reduces false positives and ensures that if threats to health or life are identified, the schools can be notified immediately by phone call.

As AI advances, so do we

The digital landscape is always evolving, and to keep up we must evolve with it. Constant innovation enables Smoothwall to continue to provide schools with best-in-class safeguarding solutions.

Our products are continually being improved - including upcoming features designed to make AI tools safer to use.

Smoothwall Filter and **Monitor** work in tandem to establish advanced protection from the risks posed by AI tools.

Schools that combine these technologies with risk assessments and clear AI policies are in the strongest position to leverage the benefits of generative AI whilst minimising the impact on student safety.