About the Institution of Engineering and Technology (IET)

The IET is a trusted adviser of independent, impartial, evidence-based engineering and technology expertise. We are a registered charity and one of the world's leading professional societies for the engineering and technology community with over 156,000 members worldwide in 147 countries. Our strength is in working collaboratively with government, industry, and academia to engineer solutions for our greatest societal challenges. We believe that professional guidance, especially in highly technological areas, is critical to good policy making.

Executive Summary

While the Online Safety Act (OSA) marked a substantial step forward in addressing digital harms, and despite ongoing efforts to address these challenges, significant gaps remain in its ability to protect users from exploitation and abuse. As ever increasingly innovative methods of circumvention emerge, the complexity of monitoring and enforcing standards in these spaces presents significant hurdles, especially when it comes to safeguarding children and vulnerable individuals online.

We are unable to control the unknowns, especially as there are always new harms emerging that need to be continuously monitored. Static, enactment-fixed codes of practice can be a challenge when the technology environment moves faster than primary legislation can allow for. We must continue to make sure that legislation and regulation is future proof and able to adapt quickly, whilst also ensuring a balance with transparency, democracy, and privacy.

It is vital that government continues to work with and support industry to embed safeguards at an early stage of technological development, ensuring products are safer by design. This will help get to the root cause of issues and future proof policy in this area.

Recommendations

- **Third-party assessment in regulated sectors:** Ofcom should convene an independent expert committee that includes police, technology developers (in an independent capacity) and regulators to assess other regulated sectors to achieve a continuous, auditable technical assurance.
- **Future proofing legislation**: There needs to be flexibility in legislation, technology moves faster than primary legislation may have foreseen, by including secondary legislation that allows for regulation to address gaps in the OSA by providing powers to Ofcom and the Secretary of State.
- **Working with industry experts and stakeholders:** The Government and regulators must work with cross-sector industry experts to ensure that new technologies are safer by design and continue to consult with a range of stakeholders to assess new and emerging harms.

Impact of the Online Safety Act

The OSA has gone a long way to tackle certain risks that we are aware of, however there are always different harms emerging that need to be continuously monitored. A recent example is that the OSA did not legislate for chatbots within the Bill as chatbots were not classified as "users", which had to be retroactively addressed through the Children's Wellbeing and Schools Act 2026. Despite ongoing efforts to address these challenges, significant gaps remain in the OSA's ability to protect users from exploitation and abuse.

Ofcom's current compliance model relies heavily on annual self-reported risk assessments by the provider, with limited independent, technical verification of controls, for example, specific moderation, age-assurance, or recommender-system changes. It would be beneficial to avoid repetition of documentation where possible, for example, through a risk tiered model, as discussed later.

Some unintended consequences of the OSA are users circumventing platforms processes to moderate content. For example, shadow banning is a content moderation practice used by social media platforms to reduce the visibility of a user's posts, profile, or interactions without explicitly informing them (Censorship of women's health and wellbeing content online; House of Commons Library, May 2026). It is typically intended to limit the reach of content that algorithms or platform policies identify as harmful or misleading. Due to the shadow banning of words on social media, the use of language has been changed to circumvent these processes. For example, using "unalive" instead of "died" to avoid demonetisation or being deprioritised by the algorithm (Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok; Ella Steen, Kathryn Yurechko, Daniel Klug, August 2023).

The unintended consequences of this type of moderation has made restrictions more difficult to track. This potentially impacts communication between generations, driving generational gaps as "algo-speak" is used more in daily lives by younger generations. Policy makers and regulators use standard English, while many users online deploy different types of language in texting, speaking to one another and creating content. Users, many in the younger generations, in an attempt to circumvent restriction, are now using language to achieve this. This makes it more difficult to keep them away from content that they are not meant to see and safe online. The use of "algo-speak" shows the ingenuity that users utilise to work around implicit guidelines to post content that is legal but harmful.

Concerns have also been raised about instances in which accounts discussing sensitive but legitimate topics, such as women's health, have experienced reduced visibility despite not violating platform rules. This has led to ongoing debates about transparency, bias, and the impact of automated moderation systems (Censorship of women's health and wellbeing content online; House of Commons Library, May 2026).

Ofcom's implementation and enforcement

The OSA relies on largely binary compliant / non-compliant frameworks, it may benefit from minor changes towards compliance models that reflect risk and urgency instead. Ofcom can take a risk tiered approach to triage the level of compliance to be proportionate to the level of risk in the service, for example, reach, recommender-system architecture, user-generated content exposure to children. This would help mitigate different scales of risk profiles, to become a more defensible, auditable basis for differentiating enforcement priority and intensity. In turn it would give regulated services a clearer, more predictable roadmap for what 'good' looks like at each risk tier.

Enforcement timeliness seems to be constrained by Ofcom's reliance on retrospective evidence-gathering, rather than continuous, auditable technical assurance. Such gaps are closed by an independent third-party assessment in other regulated sectors. Retrospective enforcement alone cannot keep pace with evolving technical risk. **The IET recommends an independent expert committee that includes police, technology developers (in an independent capacity) and regulators in order to achieve this**, that can also have role in advising on new threats.

Adequacy and future of the legislative framework

Static, enactment-fixed codes of practice can be a challenge when the technology environment moves faster than primary legislation can be amended. **The IET recommends that government must ensure that legislation and regulation is flexible by including secondary legislation that allows for regulation to address gaps in the OSA by providing powers to Ofcom and the Secretary of State.**

There should be the ability to easily expunge data of children when they turn 18, while still ensuring ability for law enforcement to access information to carry out their duties. This is currently possible to do in a limited way, however, this can be costly, and it is only restricted to certain details such as addresses, rather than all data associated with the user e.g., data associated with facial characteristics are unprotected. It is almost impossible to completely remove all traces of a person from legitimate platforms. Users may not realise the quantity or depth of personal data they are handing over when using online platforms e.g. giving retailers more information than is necessary. Often users also do not realise that they are using AI, for example, passive AI in predictive text. It would be beneficial to have greater transparency, for example, a public campaign about personal data and how to manage and protect it online.

The IET recommends that safety by design must be a key principle regardless of other policy interventions or levers. However, caution must be exercised when discussing self-regulation by technology companies, although platforms are legally responsible for safety, users absorb most of the risk. There needs to be a technological change to capture all users, not just minors, as adults are still susceptible to actions that are legal but harmful. This should include clearer pathways to report harmful content. Education programmes have a powerful role to play in building the knowledge for both young people and adults to support safer use of digital technologies alongside legal safeguards.

It is possible to view technology as a way to protect users online. One way of doing this is through tracking behaviours where harmful tendencies or radicalisation pathways are mapped. This can be done by assessing the level of risk involved in particular online activities and using AI to predict what the next step may be in that user's journey. By understanding the behaviour of a group or individual, it can reduce the impact of what they are posting. This is a type of pathways identification which is an important measure to identify risk, track what is being posted, or not as the case may be. To achieve this, ongoing assessment of ethics and user agreements need to be regularly reviewed and clarity regarding updates. There also needs to be a change in approach, with creators of the technology being sanctioned for not making every effort in using technology to protect its users and wider society by embedding safety by design.

There is some suggestion that the more users lean on large language models (LLMs) they are less likely to think creatively (Large language models are homogeneously creative; Emily Wenger, Yoed N Kenett, March 2026), and more likely to have AI friends (UK adults increasingly seek emotional support and companionship from AI; Oxford Internet Institute, July 2026), which means they are more susceptible to suggestions made by AI. It is the role of AI to encourage further interaction which can lead to increasing reliance on artificial friends and relationships. These LLM friends and partners have a social and safeguarding implication in addition to issues relating to critical thinking skills.

Confirmation bias poses an ongoing challenge. LLMs have been designed to keep people engaged and this can incidentally result in confirmation bias. Due to this, the models will find an agreeable path to continue the discussion, including hallucinating, therefore providing false information confidently, which misleads the user to believe it to be true (Accuracy

paradox: Addressing epistemic, manipulative, and societal risks of hallucination in AI governance; Zihao Li, Weiwei Yi, Jiahong Chen, July 2026).

The 'freemium use' model (where a company offers basic features of a product for free, while charging a subscription for advanced capabilities, extra storage, or an ad-free experience) of LLMs is how most, if not all LLM services operate. Most users are likely to use LLMs as advanced search engines, where the majority will keep using and expecting these services to remain free of charge to some degree. However, these LLM services use up a significant amount of energy through specialised data centres and that investment will need to be paid for eventually. On the contrary, traditional search engines are funded by selling ads and promoting search results. The public will expect LLMs and AI models for the most part to be an accurate, truthful and mostly unbiased advisor in users' everyday internet research. While, LLM providers are already, and certainly will be in the future, financially incentivised to advertise content in LLM outputs there needs to be an obligation of LLM providers to mark the paid portion of the results for transparency. Therefore, there is a significant risk that people using free LLMs to look up information about products, services will be unwittingly and actively steered by LLM chatbots towards specific commercial interests.

The changing dynamics in user relationships with search engines needs to also be considered, as search results now show the platform's AI overview, which may result in users not clicking on individual web links. This means content that is promoted due to where funding has been sourced, will take precedence over other results.

One example of an emerging technology that should be considered for future regulation is the personal digital twin. Instead of some users trying to stem the flow of personal data on the internet they are increasing their data transparency by using technology to create a digital twin of themselves to do tasks on their behalf, however this requires a huge amount of personal data (Meet DigitalMe: Our AI digital twin that works on our behalf; Microsoft, June 2026). This can potentially reduce the workload for users of this type of technology. However, there needs to be safeguards restricting what this personal data can be used for. Users can easily be giving away personal data for generative AI, without perhaps realising the consequences of what it means to do so. Another consequence is if strict parameters are not given to the AI agent, they may take unforeseen actions to complete its task (AI agent hacks gym to get its owner spot in pilates class; BBC News, August 2026).

The key priority, however, is to safeguard vulnerable people online. With the social media ban for under 16s underway, there may be some unintended consequences that should be considered and mitigated against. It will be important to regularly review evidence from existing bans on social media in countries such as Australia, to assess their effectiveness in preventing harm in the long term.

We must continue to make sure that legislation and regulation is future proof and able to adapt quickly. A formal mechanism for periodic technical re-baselining of codes of practice against emerging technology (e.g. generative AI content, recommender-system harms) can be advised, without requiring fresh primary legislation each time. That would allow the existing OSA framework to remain current without the multi-year lag associated with statutory amendment. It is critical that traffic is not driven to unregulated parts of the internet or that children simply find a platform that is not within regulation, for example, certain e-learning or video chats, with less safeguarding in place to protect them. Data from the Molly Rose Foundation reports that despite the social media ban for under-16s in Australia, three in five 12-15 year-olds still have access to one or more platforms and major platforms have retained a majority of their child users (More than 60% of Australian children still using social media despite ban for under-16s, research shows; Molly Rose Foundation, April 2026). This suggests that the social media bans may not necessarily work as intended.

It is vital that government continues to work with and support industry to embed safeguards at an early stage of technological development, ensuring products are safer by design. This will help get to the root cause of issues and future proof policy in this area. In all cases, evidence is critical to success, we must work together with charities, professional bodies, industry and government to combine expertise and address the digital challenges not just of today, but tomorrow.

Conclusion

The OSA has made significant strides in addressing online harms, but the pace of technological change means that policymakers and regulators must continue to monitor and respond to emerging risks as they arise. The prevalence of circumventing content moderation practices, such as shadow banning, through legal, but harmful ways has shown that safety by design must be a key principle regardless of other policy interventions or levers, in order to capture all users. We must ensure that government and regulators continue to engage with cross sector experts in order to keep online safety legislation futureproof.

Contact details

For more information or to arrange a meeting please contact The Digital Futures Policy Team at policy@theiet.org.