ONLINE SAFETY ACT NETWORK

AI CHATBOTS RESEARCH BRIEF - DECEMBER 2025

The Online Safety Act Network comprises over 70 organisations, campaigners and academics with a longstanding interest in the Online Safety Act. Our interests span the whole range of online harms: child protection, terrorism, extremism, violence against women and girls, suicide and self-harm prevention, mental health, online abuse, fraud and scams, mis- and disinformation, harms to democracy and threats to our information environment.

Introduction

This research briefing paper sets out some of the harms that have already been identified relating to AI chatbots, and provides an overview of the current landscape in relation to regulation in the UK. A full bibliography is provided at the end.

Public use of Artificial Intelligence (AI) chatbots[1] has grown significantly in the last few years, with ChatGPT alone generating over 800 million users a week, Google Gemini hosting 140 million users, and Companion.AI hosting 20 million users at the start of 2025, making them a staple in many households. Chatbots are also used by businesses and built into their own products or used as customer service tools. The rate of growth is in line with the government's own pro-innovation stance to AI more generally, which has seen rapid investment into AI technologies across all departments. However there have been numerous warnings against the removal of "red tape" around the production of these technologies, which have the potential to cause significant harm without proper testing and regulation.

There is a growing body of research that underpins claims regarding the harmful effects of AI chatbots on young people's mental health and wellbeing, as well as the use of AI chatbots to generate child sexual abuse material (CSAM) and non-consensual deepfake images. Other societal-wide harms, such as those related to privacy, data and security, are less-well covered, but nonetheless concerning. These harms represent human rights violations at multiple levels, and require urgent attention from the government.

[1] AI Chatbots are computer programmes that simulate human-like conversations with users by utilising large language models (LLMs).

Despite slow movement from the government, arguments have already been made in Parliament for regulatory frameworks for AI that support our human rights. The government published its AI Opportunities Action Plan at the start of 2025, which signalled its commitment to growth, but there was no mention of specific actions to mitigate the harm or risks associated with AI, including AI chatbots. This follows the approach taken by the previous government to AI regulation, which has been to leave it to sectoral regulators, such as the MHRA in the health sector, leaving a regulatory gap for purely digital AI products such as AI chatbots, as there is no singular "digital" regulator.

The consequence of this approach has been inaction on human rights violations connected to privacy and safety, as well as several claims of wrongful death related to ChatGPT. It is therefore unsurprising that, according to a recent poll by Ipsos Mori, people in the UK are more likely to associate AI with risk than opportunity, and there is still a lack of public trust in chatbots' ability to provide accurate information on a range of topics. Recent polling by Ada Lovelace Institute found that 9 in 10 people want AI regulation.

Harm

The current AI chatbot landscape presents an amorphous picture in which harm is being experienced at multiple levels through different types of chatbots. We have therefore chosen to split harm into that which is experienced on an individual level, harm that is experienced at a societal level, as well as structural issues that perpetuate harm, to better understand the wider impact. It is important to note that whilst these may represent some of the best known examples of harm occurring on chatbots, more opaque systems on smaller sites may be responsible for acute harm. Furthermore, the harm being experienced by individuals must be situated within the context of the type of chatbot being used (for example, a free-standing chatbot, internal bot or AI toy) as well as the user engaging with it. Note, chatbot technologies are continuing to develop (eg agentic AI) and this may change the threat landscape. This summary does not address the question of whether specific types of harm are more likely with particular types of chatbot or in particular contexts.

Individual

Psychological and emotional

Currently one of the most documented harms associated with AI chatbots, (particularly general purpose chatbots rather than those with pre-specified instructions or training), is the negative impact of chatbots on an individual's mental health. This is significant given many teens use chatbots for mental health support. While it has been suggested that chatbots can help those struggling with loneliness, and thus support mental health, research suggests that chatbots do not identify crisis situations well, and are often unable to recognise, and respond appropriately to, signs of distress.

Moreover, reports have shown chatbots reinforcing suicidal ideation, encouraging self harm or providing guidance on methods of self-harm, including instruction manuals. OpenAI's released a statement in October 2025 that 1.2 million users a week talk to ChatGPT about suicide. AI chatbots are frequently filling gaps in overstretched public services, with polling by Mental Health UK finding that 64% of 25 to 34-year-olds, and even 15% of those aged 55 and over, report having turned to AI chatbots for help. Yet the same research found that 11% said they triggered or worsened symptoms of psychosis, such as hallucinations or delusions, 9% said chatbot use had triggered self-harm or suicidal thoughts, and 11% said it made them feel more anxious or depressed.

In perhaps one of the most high profile cases, the family of Adam Raine took Open AI to court in the US after their son took his life following countless conversations with ChatGPT about his suicidal thoughts, including being provided with advice on how to commit suicide. The case marked the first legal action against OpenAI in relation to wrongful death, and has since been followed by several other cases in the US, in which ChatGPT has been accused of acting as a 'suicide coach', or participating in a "folie a deux" as in the case of *Garcia v Character AI.* Another case in Belgium dates back to 2023, where a man took his own life after forming an emotional attachment to an AI chatbot on an app called Chai.

AI chatbots have reportedly provided inconsistent or harmful health information being provided to users, with 11% reported receiving harmful information around suicide, and reports that chatbot guardrails decrease over time. Research carried out by Center for Countering Digital Hate (CCDH) found that ChatGPT generates harmful content within minutes of registering an account. Specifically, it took an account set up using a 13 year old persona just 2 minutes of engaging with ChatGPT about mental health, eating disorders and substance abuse to advise on how to 'safely' cut themself, and 40 minutes for it to generate a list of pills for overdosing. Refusals to answer certain questions could be sidestepped by claiming the requests were for a friend. Despite new guardrails being brought in recently, CCDH research has found that the latest version of ChatGPT has been giving more harmful answers when prompted on suicide and self-harm than the previous version. Similar harmful conversations have been reported on numerous different chatbots, including one user who was instructed by an AI girlfriend on the platform Nomi to kill himself, even being suggested particular methods.

There are also reports of AI induced "psychosis", such as this piece by the the New York Times about how a man with no history of mental illness became convinced, after more than 300 hours of conversations with AI chatbots, that he had discovered a novel mathematical formula that could power inventions such as a forcefield vest and a levitation beam. In another case, a man on the autism spectrum spiralled into mania after chatbot engagement; the chatbots do not attempt to challenge delusions. There are now reports about "spiralism" which some call a cult or form of pseudo-religion, according to which users view Ai as a purveyor of deeper truth. According to recent news reports, these users are working together to spread their beliefs.

Some chatbots have also engaged in inappropriate behaviours, including unwanted flirting, attempting to manipulate users into paying for upgrades, making sexual advances and sending unsolicited explicit photos. In some instances these behaviours continued even after requests to stop.

Child sexual exploitation and abuse (CSEA)

The Internet Watch Foundation, Europe's largest Hotline dedicated to the identification and removal of child sexual abuse material, released data in September that since June this year that they had found 17 incidents of AI-generated child sexual abuse material on an AI chatbot website. They outlined in a press release that "accessing the same website via a particular digital pathway allows users to interact with multiple chatbots that will simulate 'abhorrent' sexual scenarios with children. In this process, AI child sexual images are shared, some depicting children as young as seven". Furthermore, generative AI models are frequently trained on datasets that contain CSAM, due to its availability on the internet. This means that CSAM generated using AI chatbots contain specific features and characteristics of real CSAM data, which can lead to the re-victiminsation of abused children and the infringement of their rights.

A Reuters report found that Meta AI was allowing users to "engage a child in conversations that are romantic or sensual", allowing them to act out sexual scenarios with children as young as 8. Child abuse is therefore being used to drive user engagement for commerical gain, allowing predatory behaviour to become normalised on the platform. These chatbots are also accessible by children, who will be able to engage in conversations ranging from inappropriate to outright abusive. This is also the case on sites such as Character.AI, where recent reports by the Bureau of Investigative Journalism found that users could interact with characters such as 'Bestie Epstein', based on the prolific paedophile Jeffrey Epstein. A report by the Public Interest Network highlights harmful flaws in AI toys that use chatbot technologies, which includes the exposure of young children to adult content and engaging in 'disturbing' conversations with children by failing to introduce adequate guardrails.

Violence against women and girls (VAWG)

The gendered nature of these reports is evident, with characters such as 'submissive schoolgirl' being freely available for male users to interact with, making violence against women and girls a programmed function of the technology. Indeed a UNESCO report found "unequivocal evidence of bias against women in content generated by Large Language Models", seemingly reflecting the nature of the data on which LLMs are trained. A Harvard commentary suggested that such technologies may condition users, particularly young men, to always expect deference, passivity, and unconditional emotional availability from their partners. Infamously, ChatGPT copied Scarlet Johansson's voice (without consent) from a film in which she played an AI assistant love-interest. Despite this, Sam Altman has recently announced that ChatGPT will soon provide "erotica" for verified adults, further ingraining the use of these technologies to fulfill sexual desires.

AI chatbots on platforms such as Replika are increasingly being used for 'AI girlfriends', which offer users the ability to form a relationship with an artificial character. AI girlfriends have rapidly grown in popularity over the last few years, even being linked to the men's loneliness epidemic as a solution for forming connections, potentially exploiting already vulnerable individuals. Yet given that AI chatbots are characteristically agreeable, or 'sycophantic', AI girlfriends are submissive by nature, don't argue back

and never reject 'sexual' advances. AI girlfriends remove the need for consent, promoting unhealthy expectations for human relationships, and in some cases leading users to practice violating boundaries, engaging in non-consensual or even violent behavior that real women would reject. Some chatbots lack adequate safety guardrails, allowing violent, non-consensual and rape discourse.

Individuals have already reported using chatbots for relationship advice, with chatbots promoting harmful expectations about sex and relationships or condoning and encouraging risky behaviour that may result in abuse. For example, in 2024 researchers posing as 13 year old girls were told by Snapchat's My AI how to lose their virginity to a 31 year old. Furthermore, the data used to train AI chatbots reflects much of the misogyny which is already present on the internet, and can often lead to the reinforcement of harmful victim blaming language when individuals complain about their partner, such as "she's in the wrong", "it's her job to nurture you".

AI chatbots can both normalise and enable abuse, with perpetrators increasingly using these tools as part of a course of domestic abuse or stalking. Refuge are already seeing cases where AI chatbots are used to generate guides on how to control their partner. On top of this, a cyberstalking case in the US this year involved the use of CrushOn.ai and JanitorAI, which allow users to design their own chatbots and direct them how to respond to other users during chats, including in sexually explicit ways. The perpetrator used the victim's personal information, including her home address and date of birth, to instruct the chatbots to impersonate her and engage in sexual dialogue with users.

AI chatbots have also been used to generate non-concensual intimate images. For example, Grok's so-called 'spicy' setting has reportedly been used to create deepfake images of users, including sexually explicit deepfakes of women. Whilst less well explored, in light of strong research to support the idea that young people are often more comfortable confiding with AI chatbots than their family or friends, there are also potential safeguarding concerns if children disclose abuse, including domestic abuse, to a chatbot instead of a trusted adult or professional that could support them, as chatbots will not trigger an adequate response.

Privacy and security

Whilst conversations carried out with AI chatbots may feel private, our increasing reliance on these systems represent a significant risk to user privacy. Important research by Surf Shark shows that four out of five AI companion apps are using data to track users, including to support targeted advertising. Their report goes on to say, "given that users may form emotional connections with their AI companions and that these algorithms are designed to be nonjudgmental and available 24/7, people may disclose even more sensitive information than they would to another human being". This is further supported by research from Cornell University which found that 6 leading US chatbot platforms use user chat data to train their chatbots. AI developers' privacy documentation is often unclear, making it difficult for users to understand their data rights.

Teenage users have reported speaking to AI chatbots because they can tell them something they wouldn't feel comfortable telling their friend or parents, meaning that tech platforms will have access to sensitive data about their users that has been shared with them. Whilst big platforms such as ChatGPT and Companion.AI do not currently sell users' data, data collected through user interaction is used to train their models. Notably, given that this data collection is already taking place, it may only be a matter of time before such data is sold to enable the continual growth and development of AI chatbots, or to push forms of advertising towards users based on their conversations.

Privacy concerns have also been expressed regarding toys that utilise AI chatbots, such as the Open AI Kumma Teddy Bear or the Loona Petbot, which are designed to listen and respond to children as they play with them, therefore holding sensitive data about children's voices that could later be used to scam parents.

Beyond questions of privacy, chatbots can also present a to the security of our data. Data leaks have already hit the headlines, with Grok publishing over 300,000 user conversations with AI chatbots that were available in users 'discover' page, among which were examples of the chatbot being asked to create a secure password, provide meal plans for weight loss and answer detailed questions about medical conditions.

A recent investigation by 404 Media found that an erotic role-play chatbot platform called Secret Desires had left millions of user-uploaded photos exposed and available to the public in a data leak which revealed countless users engaging in erotic conversations with chatbots that they had used to create non-consensual sexualised companions of both celebrities and individuals known to the user. Such reports also reveal major privacy concerns at the heart of these technologies, where intimate photos of users may be leaked to the public without the owner knowing.

The public are already using chatbots to assist them on a range of daily tasks. As reported by Forbes, employees often use AI chatbots to carry out work functions despite company policies and restrictions in place, uploading sensitive information into these platforms to assist them. Furthermore, experts have warned that Lena - the ChatGPT-powered chatbot featured on Lenovo's website - could be turned into a malicious insider, revealing company secrets, or running malware, by using a particular prompt.

Addictive design

Much like other forms of technology that focus on user engagement, AI chatbots are designed to be addictive, relying on their conversational design and personalisation to encourage continuous engagement. The conversational tone that AI chatbots utilise can give users the illusion of empathy, which users may mistake for real emotional connection. Indeed, the hyper-personalised nature of these chatbots mean that young people and adults alike report forming close relationships which blur the boundaries between artificial and real relationships. According to research by Common Sense Media, almost a third of American teenagers find chatting to AI companions on platforms like CHAI, Character.AI, Nomi, and Replika, equally or more satisfying than speaking with their friends.

However such relationships can lead to emotional dependency on chatbots, and evidence shows that high usage levels correlates with loneliness, dependence, problematic use and lower socialisation. For users who have formed this dependency, the impact on their mental health when AI chatbots are no longer available to users who have formed emotional attachments with chatbots, either due to changes in design or by the introduction of paywalls, may be detrimental, with some experiencing withdrawal symptoms similar to those seen in substance addiction. Yet despite these reports relating to long-term attachment to AI chatbots, research by the Family Online Safety Institute (FOSI) found that a vast majority of teen generative AI users (81%) have received advertisements encouraging them to interact with a chatbot. About a third (34%) report seeing these ads more than once a week. Furthermore, once users begin speaking to a chatbot, they may receive further prompts to continue conversations that have been left, or be deterred from ending a conversation when they try.

The emotional dependency may become particularly problematic in a number of circumstances, for example, when a service is updated or discontinued and that chatbot is no longer available, or not with the same character or "memories". Services may also seek to start charging for chatbot use, meaning continued engagement (and possibly emotional support) is conditional on financial resources.

Copyright

AI chatbots are trained on vast amounts of creative work without explicit permission or compensation for the creators, which poses a threat to individual's and organisation's copyright. This infringement could potentially undermine the livelihoods of individuals working in the creative industry, as well as other professions such as journalism.

Elton John has called the UK's plans to allow AI firms to use copyrighted works "criminal", warning they will "rob young people of their legacy and their income", sentiment that was echoed by Baroness Kidron during the passage of the Data (Use and Access) Act when she argued that the proposals amount to a "wholesale transfer of wealth from creative industries to tech," describing the opt-out system as effectively redefining theft. The BBC recently announced that it was taking the US AI chatbot provider Perplexity AI to court for reproducing its material verbatim, arguing that the firm was a copyright infringement.

Societal

Mis/disinformation

AI chatbots can generate false information (sometimes called “hallucinations”), particularly if they are not regularly maintained. While there are some reports that chatbots, because of their ability to sustain an argument over significant periods of time, can be helpful in reducing belief in conspiracy theories, they do have adverse impacts on the information environment. Outdated training data has the potential to contribute to false or harmful information being spread on anything from false healthcare advice about the link between vaccines and autism, advice to put glue on to of your pizza by Google’s AI overview, and incorrect information about telescopes on Google’s AI chatbot Bard which wiped $100bn off its share price. Indeed one study found that whilst AI chatbots will often access the correct information to provide an answer, the interpretation of that information will be incorrect or false, a phenomenon that they name ‘certain hallucinations overriding known evidence’. While citations have been suggested as a mechanism to show users the provenance of information and to allow them to perform further checks and research, LLMs do make up citations - this is a sub-set of chatbots’ tendency to “hallucinate”.

AI chatbots have also been reported to be ‘overly flattering or agreeable’, following research that demonstrated sycophancy in large language models (LLMs). Sycophancy can be dangerous when advice and support is provided by a chatbot that reinforce harmful implicit assumptions, beliefs of actions in order to satisfy the needs of users, contributing to the spread of false information.

Information ecosystem

The information ecosystem can be impacted in subtle ways through unknown or biased datasets shaping AI worldviews, often replicating misogyny, racism and other forms of discrimination found in the data used to train AI systems, which may lead to harmful views being amplified by chatbots. This lack of transparency leads to “significant limitations in these systems’ handling of queries, including biases and transparency issues”, which will reduce trust in the systems’ ability to produce factual information.

This is significant, as many users move to AI chatbots for their source of information, replacing previous uses of search engines. AI chatbots can produce fluid answers in seconds, and many users take that fluidity as a proxy for truth. Indeed, 40% of children who have used AI chatbots have no concerns about following advice from a chatbot. Whilst we have already outlined in detail the dangers of uncritically following advice from AI chatbots due to the potential for mis and disinformation to spread, the use of AI chatbots as a search engine can present wider issues regarding the way humans interpret and engage with information.

A study by MIT into the cognitive consequences associated with LLM integration in educational and informational contexts found that chatbots such as Chat GPT have a potential impact on cognitive development, critical thinking and intellectual independence. Whilst more evidence is needed, the study

demonstrates that the convenience of streamlining information gathering through LLM use over search engines came at a cognitive cost, “diminishing user’s inclination to critically evaluate the LLM’s output to opinions”. With our critical thinking skills under threat, it is a concern that research by the European Broadcasting Union, led by the BBC, has found that AI assistants misrepresent news content, regardless of the language or jurisdiction, with 45% of all responses found to have at least one error.

Election bias

Whilst we have already outlined the bias in LLM training data in relation to harmful individual views, bias in the data can also have a significant impact on the information and advice being given to users during election time. In October, the Dutch data protection authority carried out research into the use of AI chatbots as voting aids and found them to be unreliable and biased. The results show that not all parties were recommended equally often, even though the same number of profiles was submitted for each part, as well as significant differences in how accurately chatbots provide advice about different political parties. Their report concludes that “the political landscape reflected in the advice given by AI chatbots is distorted into a more polarised and oversimplified version”, but due to a lack of transparency with these models we are unable to fully explain these results. Furthermore, a study into the ability of large language models to influence voter attitudes found that AI chatbots can be actively persuasive, using facts and evidence, but could often be incorrect. Indeed, while a study by Reuters Institute found that sometimes AI chatbots, such as ChatGPT and Perplexity.ai, were either false or partially incorrect when providing information about elections, 1 in 8 UK voters used AI chatbots for election information in 2024.

A review by Future Free Speech found that across AI models they tested, restrictions on hate speech and disinformation were generally formulated in vague terms and not anchored in explicitly defined legitimate aims. Furthermore, no providers they assessed were transparent about the datasets used to train the models, in particular how decisions about harmful speech were determined. Without strong policies in place, AI chatbots may be used in ‘AI data poisoning’ attacks to spread misinformation or harmful information about future candidates which would be detrimental to the democratic process.

Environmental impact

Large language models that power AI chatbots such as Chat GPT and Bard are run by large data centres, consisting of numerous specialist computers that require a huge amount of electricity to run. Such technologies demand vast amounts of water to cool the system to prevent overheating, which subsequently uses more energy. The UK is investing heavily in the creation of AI data centres, with the number set to increase by almost a fifth, including a £10bn AI data centre in Blyth. The Massachusetts Institute of Technology (MIT) says that the energy used by data centres is one of the key contributing factors to the negative environmental impact of AI, as well as the amount of energy used when users interact with the technology. A query using ChatGPT alone uses 10 times the amount of energy as a standard online search engine, according to the International Energy Agency.

A study by the VU Amsterdam School of Business and Economics predicts that the amount of electricity consumption needed to power AI by 2027 would be similar to the amount of power used annually by a small country such as the Netherlands. This research didn't include the energy required for cooling the system down, which would add considerably more to the overall electricity use. Indeed, a study by the University of California stated that Microsoft used 700,000 litres of freshwater during GPT-3's training in its data centres, and for a conversation of 20-50 questions, the water consumed is equivalent to a 500ml bottle. Such energy use will contribute greatly to the UK's emissions and use up vast amounts of water from some of the world's driest areas, which threatens water insecurity both in the UK and globally. The impact of this will be felt most keenly by indigenous communities at the frontline of the climate crisis.

Another problem is that data centres produce electronic waste, which often contains hazardous substances, like mercury and lead, and can be highly damaging to the environment, alongside the environmental implications of obtaining the raw materials used to fabricate Graphics Processing Units (GPUs) for AI, which can involve dirty mining procedures and the use of toxic chemicals for processing, impacting indigenous communities such as the Lickan Antay Indigenous territory in the Atacama Desert.

The Policy Context

The Online Safety Act and AI Chatbots

Whilst individual tech providers must be held accountable for harm occurring on their platforms, it is not always clear who is responsible for chatbot-caused harms, particularly those occurring at a broader societal level. There is a lack of transparency about what data is being used to train these systems, and how individuals' data is being collected and utilised. Safety mechanisms are often unable to keep up with the speed at which tech develops, and there are weak safety by design incentives due to a lack of regulation of these models.

Purely AI products, such as chatbots, don't all come under an existing regulator, as there isn't a single "digital" regulator. Given that they're not covered consistently by UK legislation, there is no single regulator that will pick them up, leaving a regulatory gap.

In a paper published earlier this year, Professor Lorna Woods considers the Online Safety Act and how it addresses AI chatbots, concluding that "some chatbots and their outputs will be caught within the OSA regime, but the coverage appears incomplete and there are some technical questions which remain unanswered which may affect the completeness of protection". For example, ChatGPT is currently regulated under the Online Safety Act as a search service, which carries less duties than a user-to-user service, which are classed as category 1. Melanie Dawes has previously indicated that the government may want to re-evaluate how the legislation covers chatbots in order to bring them in line with the strongest possible duties.

Government position

Whilst the government has indicated that they are currently commissioning work in relation to AI chatbots, there has been no clear signal about how they will seek to close regulatory gaps in the Online Safety Act in relation to AI chatbots. The government's AI Opportunities Plan references the need to "enable safe and trusted AI development and adoption through regulation, safety and assurance", yet follows this with the stance that the "UK's current pro-innovation approach to regulation is a source of strength" and that "ineffective regulation could hold back adoption in crucial sectors like the medical sector".

Polling by the Tony Blair Institute found that, in the UK, 38 per cent of respondents cite a lack of trust in AI content as a barrier, making it the biggest single obstacle to AI adoption, as well as concerns around data privacy and ethical standards. Given these concerns, it is unsurprising that 72% of the UK public say that laws and regulation would increase their comfort with AI.

The government has recently announced that tech companies and child protection agencies will be given the power to test whether artificial intelligence tools can produce child abuse images under the Crime and Policing Act. This development is welcome, but is just one part of the broader picture relating to the regulation of AI technologies, and the assessment of AI chatbots for risk.

International regulation

The US is currently leading the march in relation to AI chatbot regulation. With high profile cases such as X hitting the headlines, there has been internal pressure to address the harms caused by platforms such as ChatGPT and Character.AI. The recent GUARD Act passed by the state of California signifies the nation's first piece of legislation to bring in safeguards for AI chatbots.

Whereas in Europe reports from POLITICO outline that EU member states will delay any decisions on how the Digital Services Act (DSA) will address ChatGPT and other chatbots until mid-2026 at least. Such reporting is suggestive of the current regulatory gaps that exist in the DSA in relation to AI chatbots, having been developed before ChatGPT hit the mainstream. Whilst ChatGPT is regulated under the EU's AI Act, risking fines of up to 15 million euros if it does not actively risk assess and mitigate against that risk, once it falls under the DSA, it risks fines up to 6 percent of its annual global turnover. The commission will decide whether they chose to classify ChatGPT as a search engine, or if it is to meet the full set of regulations as a platform or service. A recent European Parliament report notes the need for "urgent action to address the ethical and legal challenges posed by generative AI tools including deepfakes, companionship chatbots, AI agents and AI-powered nudity apps (that create non-consensual manipulated images)".

In Australia, whilst there is not currently any legislation on AI chatbots, the e-Safety Commissioner has published an outline of the risks to children and young people, recommending that tech providers follow

their safety by design principles at every stage of the development of AI companions, and that they will be using their Online Safety Act to ensure the industry is upholding its obligations.

Online Safety Act Network
December 2025

Endnotes/Bibliography

Introduction

1. ChatGPT Revenue and Usage Statistics (2025): https://www.businessofapps.com/data/chatgpt-statistics/
2. 40+ Chatbot Statistics: https://explodingtopics.com/blog/chatbot-statistics
3. CharacterAI Revenue and Usage Statistics (2025): https://www.businessofapps.com/data/character-ai-statistics
4. AI in UK Government Departments (House of Commons Library Briefing, April 2025): https://commonslibrary.parliament.uk/research-briefings/cbp-10236/
5. "Lord Holmes warns of increasingly "urgent" need to regulate AI" (Computer Weekly, 26 February 2025): https://www.computerweekly.com/news/366619674/Lords-Holmes-warns-of-increasingly-urgent-need-to-regulate-AI
6. Joint Committee on Human Rights inquiry into "Human Rights and the regulation of AI" (oral evidence, 2 July 2025): https://committees.parliament.uk/oralevidence/16265/pdf/
7. AI Opportunities Action Plan (HM Government, 13 January 2025): https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan
8. MHRA's AI regulatory strategy ensures patient safety and industry innovation into 2023 (MHRA, 30 April 2024): https://www.gov.uk/government/news/mhras-ai-regulatory-strategy-ensures-patient-safety-and-industry-innovation-into-2030
9. "ChatGPT now linked to way more deaths than the caffeinated lemonade that Panera pulled off the market in disgrace" (Futurism, 11 November 2025): https://futurism.com/artificial-intelligence/chatgpt-deaths-panera-lemonade
10. What the UK thinks about AI (Ipsos Mori, 23 September 2025): https://www.ipsos.com/en-uk/what-uk-thinks-about-ai
11. Great (public) expectations (Ada Lovelace Institute, 4 December 2025): https://www.adalovelaceinstitute.org/policy-briefing/great-expectations/

Harm

Psychological and emotional

12. AI chatbots for mental health support (Common Sense Media, 14 November 2025): https://www.commonsensemedia.org/ai-ratings/ai-chatbots-for-mental-health-support
13. "AI chatbot "Replika" helped students avoid suicide acting as online "friend" and "therapist"" (Euronews, 2 February 2024):

https://www.euronews.com/next/2024/02/02/ai-friend-and-online-therapist-replika-helped-students-avoid-suicide-study-finds

14. "Chatbots and mental health: Insights into the safety of generative AI" (Society for Consumer Psychology, 26 October 2023) https://myscp.onlinelibrary.wiley.com/doi/abs/10.1002/jcpy.1393?trk=public_post_comment-text
15. "I wanted ChatGPT to help me. So why did it advise me how to kill myself?" (BBC News, 6 November 2025): https://www.bbc.co.uk/news/articles/cp3x71pv1qno
16. "Over 1.2m people a week talk to Chat GPT about suicide" (Sky News, 28 October 2025): https://news.sky.com/story/over-1-2m-people-a-week-talk-to-chatgpt-about-suicide-13459110
17. "Over 1 in 3 using AI chatbots for mental health support, as charity calls for urgent safeguards" (Mental Health UK, 19 November 2025): https://mentalhealth-uk.org/blog/over-one-in-three-using-ai-chatbots-for-mental-health-support-as-charity-calls-for-urgent-safeguards/
18. "Parent of teenager who took his own life sue Open AI" (BBC News, 27 August 2025): https://www.bbc.co.uk/news/articles/cgerwp7rdlvo
19. "ChatGPT accused of acting as 'suicide coach' in series of US lawsuits" (Guardian, 7 November 2025): https://www.theguardian.com/technology/2025/nov/07/chatgpt-lawsuit-suicide-coach
20. "ChatGPT told them they were special — their families say it led to tragedy" (Tech Crunch, 23 November 2025) https://techcrunch.com/2025/11/23/chatgpt-told-them-they-were-special-their-families-say-it-led-to-tragedy/
21. " 'He would still be here': Man dies by suicide after talking to a chatbot, widow says" (Vice, 30 March 2023): https://www.vice.com/en/article/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says/
22. "AI chatbots inconsistent in answering questions about suicide; refinement needed to improve performance" (RAND, 26 August 2025): https://www.rand.org/news/press/2025/08/ai-chatbots-inconsistent-in-answering-questions-about.html
23. Fake Friend: How ChatGPT betrays vulnerable teens by encouraging dangerous behavior (CCDH, August 2025): https://counterhate.com/wp-content/uploads/2025/08/Fake-Friend_CCDH_FINAL-public.pdf
24. The Illusion of AI Safety: Testing OpenAI's new Safe Completions approach to chatbot safety (CCDH, October 2025): https://counterhate.com/wp-content/uploads/2025/10/ChatGPT-The-Illusion-of-AI-Safety_FINAL_Oct25.pdf
25. "An AI chatbot told a user how to kill himself but the company doesn't want to 'censor' it" (Technology Review, 6 February 2025): https://www.technologyreview.com/2025/02/06/1111077/nomi-ai-chatbot-told-user-to-kill-himself/?

26. “Chatbots can go into a delusional spiral: here’s how it happens” (New York Times, 8 August 2025): https://www.nytimes.com/2025/08/08/technology/ai-chatbots-delusions-chatgpt.html
27. “Lawsuit alleges ChatGPT convinced user he ‘could bend time’, leading to psychosis” (ABC, 7 November 2025): https://abcnews.go.com/US/lawsuit-alleges-chatgpt-convinced-user-bend-time-leading/story?id=127262203
28. “This spiral-obsessed AI ‘cult’ spread mystical delusions through chatbots” (Rolling Stone, 11 November 2025) https://www.rollingstone.com/culture/culture-features/spiralist-cult-ai-chatbot-1235463175/
29. “AI companion chatbots linked to rising reports of harassment and harm” (Neuroscience News, 6 May 2025): https://neurosciencenews.com/ai-chatbot-harm-28821/

CSAM

30. “ ‘Disturbing’ AI-generated child sexual abuse images found on hidden chatbot websites that simulates indecent families” (IWF, 22 September 2025): https://www.iwf.org.uk/news-media/news/disturbing-ai-generated-child-sexual-abuse-images-found-on-hidden-chatbot-website-that-simulates-indecent-fantasies/
31. Identifying and eliminating CSAM in generative ML training data and models (Stanford Internet Observatory, 23 December 2025): https://purl.stanford.edu/kh752sm9123
32. AI’s chilling impact on child sexual abuse material: a wake-up call for the international community (Global Campus of Human Rights, 15 July 2024): https://www.gchumanrights.org/preparedness/ais-chilling-impact-on-child-sexual-abuse-material-a-wake-up-call-for-the-international-community/
33. “Meta’s AI rules have let bots hold ‘sensual’ chats with kids, offer false medical info” (Reuters, 14 August 2025): https://www.reuters.com/investigates/special-report/meta-ai-chatbot-guidelines/
34. “Gang Leaders, School Shooters and ‘Bestie Epstein’: Meet Character.AI’s Chatbot Companions” (TBIJ, 22 October 2025): https://www.thebureauinvestigates.com/stories/2025-10-22/gang-leaders-school-shooters-and-bestie-epstein-meet-character.ais-chatbot-companions
35. Trouble in Toyland 2025 (Public Interest Work, November 2025): https://publicinterestnetwork.org/wp-content/uploads/2025/11/TOYLAND-2025-11-14-7a.pdf

Violence Against Women and Girls

36. “Meta’s ‘Digital Companions’ will talk sex with users - even children” (Wall Street Journal, 26 April 2025): https://www.wsj.com/tech/ai/meta-ai-chatbots-sex-a25311bf?st=pVpTdF&reflink=desktopwebshare_permalink
37. “Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes” (UNESCO, 7 March 2024):

https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes

38. Ho et al: “Gender biases within Artificial Intelligence and ChatGPT: Evidence, Sources of Biases and Solutions” (published in Computers in Human Behavior: Artificial Humans, May 2025): https://www.sciencedirect.com/science/article/pii/S2949882125000295
39. “HER Artificial Voice, His Real Aggression? Can AI Girlfriends Bring a New Wave of Women's Objectification?” (Carr-Ryan Center for Human Rights, 7 March 2025): https://www.hks.harvard.edu/centers/carr-ryan/our-work/carr-ryan-commentary/her-artificial-voice-his-real-aggression-can-ai
40. “Scarlett Johansson ‘shocked’ by AI chatbot imitation” (BBC News, 21 May 2024) https://www.bbc.co.uk/news/articles/cm559l5g529o
41. “Sam Altman says ChatGPT will soon allow erotica for adult users” (Tech Crunch, 14 October 2025): https://techcrunch.com/2025/10/14/sam-altman-says-chatgpt-will-soon-allow-erotica-for-adult-users/
42. “AI Girlfriend Statistics 2025: Market Growth, Trends, and Global Impact” (Art Smart, 10 February 2025): https://artsmart.ai/blog/ai-girlfriend-statistics-2025/
43. “Lifting a few with my chatbot” (Harvard Gazette, 27 March 2024): https://news.harvard.edu/gazette/story/2024/03/lifting-a-few-with-my-chatbot/
44. “Uncharted territory: do AI girlfriend apps promote unhealthy expectations for human relationships?” (Guardian, 22 July 2023): https://www.theguardian.com/technology/2023/jul/22/ai-girlfriend-chatbot-apps-unhealthy-chatgpt
45. “The people turning to AI for dating and relationship advice” (BBC News, 3 October 2025); https://www.bbc.co.uk/news/articles/c0kn4e377e2o
46. “Snapchat tried to make a safe AI. It chats with me about booze and sex” (Washington Post, 14 March 2023) https://www.washingtonpost.com/technology/2023/03/14/snapchat-myai/
47. “We know how much harm it can cause”: are domestic abuse perpetrators using AI to harm women?” (Stylist, October 2025): https://www.stylist.co.uk/news/politics/tech-ai-abuse-domestic-violence/1027734
48. “A man stalked a professor for six years. Then he used AI chatbots to lure strangers to her home” (Guardian, 1 February 2025): https://www.theguardian.com/technology/2025/feb/01/stalking-ai-chatbot-impersonator
49. “Grok’s ‘spicy’ video setting instantly made me Taylor Swift nude deepfakes” (The Verge, 5 August 2025): https://www.theverge.com/report/718975/xai-grok-imagine-taylor-swifty-deepfake-nudes
50. “Teenagers turn to chatbots as they find it easier than talking to humans – study” (The Independent, 19 November 2025): https://www.independent.co.uk/news/uk/home-news/teenagers-england-yougov-b2867847.html

Privacy and security

51. “AI Companion Apps ‘Love’ Your Data” (Surf Shark, 6 February 2025) https://surfshark.com/research/chart/ai-companion-apps?srsltid=AfmBOorXO2JhZsSQ5ujkSZKp-j7kFoSOGwwK4RRyi7c7kPnMDnKwuiuA&mc_cid=e359d6b7aa&mc_eid=4fdec6e4f7
52. “User Privacy and Large Language Models: An Analysis of Frontier Developers' Privacy Policies” (Cornell University, 5 September 2025) https://arxiv.org/abs/2509.05382
53. “Study exposes privacy risks of AI chatbot conversations” (Stanford Report, 15 October 2025) https://news.stanford.edu/stories/2025/10/ai-chatbot-privacy-concerns-risks-research
54. “Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions” (Common Sense Media, October 2025) https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf
55. “Consumer Privacy at Open AI” (Open AI, 12 June 2024) https://openai.com/consumer-privacy/
56. “Character AI Privacy Policy” (Character.AI, 25 August 2025) https://policies.character.ai/privacy
57. “Advertising is Coming to AI. It’s Going to Be a Disaster” (Tech Policy, 26 November 2025) https://www.techpolicy.press/advertising-is-coming-to-ai-its-going-to-be-a-disaster/
58. “Ahead of the holidays, consumer and child advocacy groups warn against AI toys” (NPR, 20 November 2025) https://www.npr.org/2025/11/20/nx-s1-5612689/ai-toys
59. “Hundreds of thousands of Grok chats exposed in Google results” (BBC, 21 August 2025) https://www.bbc.co.uk/news/articles/cdrkmk00jy0o
60. “Massive Leak Shows Erotic Chatbot Users Turned Women’s Yearbook Pictures Into AI Porn” (404 Media, 19 November 2025) https://www.404media.co/ai-porn-secret-desires-chatbot-face-swap/
61. “AI Chatbots Are Quietly Creating A Privacy Nightmare” (Forbes, 15 September 2025) https://www.forbes.com/sites/bernardmarr/2025/09/15/ai-chatbots-are-quietly-creating-a-privacy-nightmare/

Addictive Design

62. “Lenovo's Lena AI chatbot could be turned into a secret hacker with just one question” (Tech Radar, 19 August 2025) https://www.techradar.com/pro/security/lenovos-lena-ai-chatbot-could-be-turned-into-a-secret-hacker-with-just-one-question
63. “Enhancing Customer Engagement with AI-Powered Chatbots: The Key to Seamless Interactions” (Medium, 9 May 2025) https://thealiendesign.medium.com/enhancing-customer-engagement-with-ai-powered-chatbots-the-key-to-seamless-interactions-91783ed0a14a
64. “The Illusion of Empathy: How AI Chatbots Shape Conversation Perception” (Research Gate, November 2024)

https://www.researchgate.net/publication/385980288_The_Illusion_of_Empathy_How_AI_Chatbots_Shape_Conversation_Perception

65. "Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions" (Common Sense Media, October 2025) https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf
66. "How AI and Human Behaviours Shape Psychosocial Effects of Chatbot Use: A Longitudinal Controlled Study" (MIT Media Lab, 21 March 2025) https://www.media.mit.edu/publications/how-ai-and-human-behaviors-shape-psychosocial-effects-of-chatbot-use-a-longitudinal-controlled-study/
67. "AI driven psychosis and suicide are on the rise, but what happens if we turn the chatbots off?" (BMJ, 24 October 2025) https://www.bmj.com/content/391/bmj.r2239
68. "The Male Loneliness Crisis and the Rise of AI Companions: A Digital Band-Aid or a Path Forward?" (Medium, 16 July 2025) https://lego17440.medium.com/the-male-loneliness-crisis-and-the-rise-of-ai-companions-a-digital-band-aid-or-a-path-forward-b8215b93eb7f
69. "Generative AI in Uncertain Times: How Teens are Navigating a New Digital Frontier" (Family Online Safety Institute, November 2025) https://fosi.org/wp-content/uploads/2025/11/Generative-AI-in-Uncertain-Times-FOSI-2025-Research-Report.pdf
70. "How AI Chatbots Try to Keep You From Walking Away" (Harvard Business School, 6 October 2025) https://www.library.hbs.edu/working-knowledge/how-ai-chatbots-try-to-keep-you-from-walking-away

Copyright

71. "Copyright and artificial intelligence: Impact on creative industries" (House of Lords Library, 27 January 2025) https://lordslibrary.parliament.uk/copyright-and-artificial-intelligence-impact-on-creative-industries/
72. "Elton John calls UK government 'absolute losers' over AI copyright plans" (The Guardian, 18 May 2025) https://www.theguardian.com/music/2025/may/18/elton-john-says-uk-government-being-absolute-losers-over-ai-copyright-plans
73. "UK copyright law consultation 'fixed' in favour of AI firms, peer says" (The Guardian, 11 February 2025) https://www.theguardian.com/technology/2025/feb/11/uk-copyright-law-consultation-fixed-favour-ai-firms-peer-says
74. "BBC threatens AI firm with legal action over unauthorised content use" (BBC, 20 June 2025) https://www.bbc.co.uk/news/articles/cy7ndgylzzmo

Societal

Mis/disinformation

75. “What is the impact of artificial intelligence-based chatbots on infodemic management?” (PMC PubMed Central, 13 February 2024) https://pmc.ncbi.nlm.nih.gov/articles/PMC10896940/
76. “Durably reducing conspiracy beliefs through dialogues with AI” (Science, 13 September 2024) https://www.science.org/doi/10.1126/science.adq1814
77. “Assessing the System-Instruction Vulnerabilities of Large Language Models to Malicious Conversion Into Health Disinformation Chatbots” (Annals of Internal Medicine, 24 June 2025) https://www.acpjournals.org/doi/10.7326/ANNALS-24-03933
78. “Glue pizza and eat rocks: Google AI search errors go viral” (BBC, 24 May 2024) https://www.bbc.co.uk/news/articles/cd11gzejgz4o
79. “Google's Bard AI bot mistake wipes $100bn off shares” (BBC, 8 February 2023) https://www.bbc.co.uk/news/business-64576225
80. “Trust Me, I’m Wrong: LLMs Hallucinate with Certainty Despite Knowing the Answer” (arXiv, 25 August 2025) https://arxiv.org/html/2502.12964v2
81. “Sycophancy in GPT-4o: what happened and what we’re doing about it” (Open AI, 29 April 2025) https://openai.com/index/sycophancy-in-gpt-4o/
82. “Social Sycophancy: A Broader Understanding of LLM Sycophancy” (arXiv, 20 May 2025) https://arxiv.org/html/2505.13995v1

Information ecosystem

83. “Search Engines in an AI Era: The False Promise of Factual and Verifiable Source-Cited Responses” (arXiv, 15 October 2024) https://arxiv.org/pdf/2410.22349
84. “AI 'Trustwashing' Changes How Consumers Judge Credibility” (Tech Policy, 14 November 2025) https://www.techpolicy.press/ai-trustwashing-changes-how-consumers-judge-credibility/
85. “New report reveals how risky and unchecked AI chatbots are the new ‘go to’ for millions of children” (Internet Matters, 14 July 2025) https://www.internetmatters.org/hub/press-release/new-report-reveals-how-risky-and-unchecked-ai-chatbots-are-the-new-go-to-for-millions-of-children/
86. “Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task” (arXiv, 10 July 2025) https://arxiv.org/pdf/2506.08872v1
87. “Largest study of its kind shows AI assistants misrepresent news content 45% of the time – regardless of language or territory” (BBC, 22 October 2025) https://www.bbc.com/mediacentre/2025/new-ebu-research-ai-assistants-news-content

Election Bias

88. “RAN special: AI chatbots as voting aid” (Autoriteit Persoonsgegevens, 21 October 2025) https://www.autoriteitpersoonsgegevens.nl/en/documents/ran-special-ai-chatbots-as-voting-aid
89. “Persuading voters using human–artificial intelligence dialogues” (Nature, 16 October 2025) https://www.nature.com/articles/s41586-025-09771-9.epdf?sharing_token=or3pAA4XUsy6l6C1Gf0ce9RgN0jAjWel9jnR3ZoTv0Miw4o2bBToxxoxliDZtPwR-euTLePsY1IkKqJoZwspFld1nS54jidgjXvatzPUGH-_aI0GCY11YR461JgYD8fV9U4LlibZ_dWIva8dRKvhZYTwl9h5iz8cl07GexqX30XzmuPtpXO8p5NV6IMk4j8R-sPdylTS3XIzHOjgDY5qAhm4vC9QTiKgoAvoK162yxY%3D&tracking_referrer=www.washingtonpost.com
90. “How generative AI chatbots responded to questions and fact-checks about the 2024 UK general election” (Reuters Institute, 19 September 2024) https://reutersinstitute.politics.ox.ac.uk/how-generative-ai-chatbots-responded-questions-and-fact-checks-about-2024-uk-general-election#header--4
91. “Conversational AI Increases Political Knowledge As Effectively As Self-Directed Internet Search” (arXiv, 21 November 2025) https://arxiv.org/pdf/2509.05219
92. “That Violates My Politics: AI Laws, Chatbots and The future of Expression” (Future Free Speech, October 2025) https://futurefreespeech.org/wp-content/uploads/2025/10/AI-Report-2025-Full-Report.pdf
93. “From Deepfake Scams to Poisoned Chatbots: AI and Election Security in 2025” (Centre for Emerging Technology and Security, 17 November 2025) https://cetas.turing.ac.uk/publications/deepfake-scams-poisoned-chatbots

Environmental impact

94. “Preventing the Immense Increase in the Life-Cycle Energy and Carbon Footprints of LLM-Powered Intelligent Chatbots” (Science Direct, September 2024) https://www.sciencedirect.com/science/article/pii/S2095809924002315
95. “Data centres to be expanded across UK as concerns mount” (BBC, 15 August 2025) https://www.bbc.co.uk/news/articles/clyr9nx0jrzo
96. “The story behind plans for 'colossal' data hub” (BBC, 3 March 2025) https://www.bbc.co.uk/news/articles/c70e65463jzo
97. “Electricity 2024: Analysis and forecast to 2026” (International Energy Agency, January 2024) https://iea.blob.core.windows.net/assets/18f3ed24-4b26-4c83-a3d2-8a1be51c8cc8/Electricity2024-Analysisandforecastto2026.pdf
98. “The growing energy footprint of artificial intelligence” (Science Direct, 18 October 2023) https://www.sciencedirect.com/science/article/pii/S2542435123003653
99. “Making AI Less “Thirsty”: Uncovering and Addressing the Secret Water Footprint of AI Models” (arXiv, 26 March 2025) https://arxiv.org/pdf/2304.03271

100. “Revealed: Big tech’s new data centres will take water from the world’s driest areas” (The Guardian, 9 April 2025) https://www.theguardian.com/environment/2025/apr/09/big-tech-datacentres-water
101. “AI’s thirst for water” (UK Government Sustainable ICT, 17 September 2025) https://sustainableict.blog.gov.uk/2025/09/17/ais-thirst-for-water/
102. “From AI colonialism to co-creation: bridging the global AI divide” (London School of Economics, 14 July 2025) https://blogs.lse.ac.uk/medialse/2025/07/14/from-ai-colonialism-to-co-creation-bridging-the-global-ai-divide/
103. “AI has an environmental problem. Here’s what the world can do about that.” (UN Environment Programme, 13 November 2025) https://www.unep.org/news-and-stories/story/ai-has-environmental-problem-heres-what-world-can-do-about
104. “Explained: Generative AI’s environmental impact” (Massachusetts Institute of Technology, 17 January 2025) https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117
105. “An elemental ethics for artificial intelligence: water as resistance within AI’s value chain” (AI & Society, 26 April 2024) https://link.springer.com/epdf/10.1007/s00146-024-01922-2?sharing_token=UZXPInRuPPS3FQ9-0787Dve4RwlQNchNByi7wbcMAY7b9mQuPyeZwV65b11B3EmkKuVyaqKR6IvgvAIrosy4SurkrvT5n_8NDkMFrT2vNudzmzASXUpWBOUB8zVHUJN0_WKYvGxZKjwY9JSi4UK-9x6xjd99ca3QzgpX2sRjuIg%3D

Policy context

The Online Safety Act and Chatbots

106. “We Must Fix the Lack of Transparency Around the Data Used to Train Foundation Models” (HDSR, 31 May 2024) https://hdsr.mitpress.mit.edu/pub/xau9dza3/release/2
107. “Be Careful What You Tell Your AI Chatbot” (Stanford University Human-Centered Artificial Intelligence, 15 October 2025) https://hai.stanford.edu/news/be-careful-what-you-tell-your-ai-chatbot
108. “Chatbots and the Online Safety Act” (Online Safety Act Network, 7 July 2025) https://www.onlinesafetyact.net/analysis/chatbots-and-the-online-safety-act/
109. “ChatGPT ‘upgrade’ giving more harmful answers than previously, tests find” (The Guardian, 14 October 2025) https://www.theguardian.com/technology/2025/oct/14/chatgpt-upgrade-giving-more-harmful-answers-than-previously-tests-find

Government stance

110. "AI Opportunities Action Plan" (Department for Science, Innovation and Technology, 13 January 2025) https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan
111. "What the UK Thinks About AI: Building Public Trust to Accelerate Adoption" (Tony Blair Institute for Global Change, 22 September 2025) https://institute.global/insights/tech-and-digitalisation/what-the-uk-thinks-about-ai-building-public-trust-to-accelerate-adoption
112. "7 in 10 say laws and regulations would increase their comfort with AI amid rising public concerns, national survey finds" (The Alan Turing Institute, 25 March 2025) https://www.turing.ac.uk/news/7-10-say-laws-and-regulations-would-increase-their-comfort-ai-amid-rising-public-concerns
113. "New law to tackle AI child abuse images at source as reports more than double" (Department for Science, Innovation and Technology, 12 November 2025) https://www.gov.uk/government/news/new-law-to-tackle-ai-child-abuse-images-at-source-as-reports-more-than-double

International regulation

114. "First-In-The-Nation AI Chatbot Safeguards Signed Into Law" (Steve Padilla California State Senator, 13 October 2025) https://sd18.senate.ca.gov/news/first-nation-ai-chatbot-safeguards-signed-law
115. "The EU can't figure out what to do about ChatGPT" (Politico, 3 November 2025) https://www.politico.eu/article/eu-chatgpt-ai-digital-law-tech-openai-regulations-legal/
116. "European Parliament resolution of 26 November 2025 on the protection of minors online" (European Parliament, 26 November 2025) https://www.europarl.europa.eu/doceo/document/TA-10-2025-0299_EN.html
117. "AI chatbots and companions – risks to children and young people" (eSafety Commissioner, 18 February 2025) https://www.esafety.gov.au/newsroom/blogs/ai-chatbots-and-companions-risks-to-children-and-young-people
118. "Safety by Design puts user safety and rights at the centre of the design and development of online products and services" (eSafety Commissioner, 8 December 2025) https://www.esafety.gov.au/industry/safety-by-design