

AI & Online Safety – Guide for Parents, Guardians, and Teachers

This guide outlines some of the risks associated with generative AI (GAI) and provides resources for parents, guardians, and teachers to best protect children.

What are the risks?

GAI tools can be beneficial for young users. For example, some tools can be used to tailor lessons or homework to the child's educational needs. Using chatbots can also be a great way to practice social interactions before they happen, which may be particularly useful for neurodivergent users. However, using GAI tools comes with risks. Some of these are discussed below.

Exposure to harmful material: AI may be used to create deepfake content - images, video, and even audio, which can be used in cases of bullying. Misinformation is often spread on social media platforms, and algorithmic biases may create an echo chamber where stereotypes are reinforced, and users radicalised. Software such as AI companions are often unmoderated and may expose users to age-inappropriate conversations and even advice on topics such as health, sex, and self-harm.

Extortion: Recent reports have identified a rise in the use of AI-generated indecent images in sextortion cases among young people. This involves gathering ordinary images of the victim, usually taken from a social media platform, and using AI to create fake explicit content. Worryingly, given advancements in technology, this content is often very realistic and may be used to extort victims even if they have never shared an indecent image of themselves. The availability of tools which enable users to remove clothes from individuals in an image, means they are unfortunately used in some cases of peer bullying.

User Dependency: Software such as AI companions are designed to encourage high engagement, which may lead to addiction and user dependency. Young users are more vulnerable to this, as they may struggle to understand the differences between interacting with AI and humans.

Real-Life Instances of AI Harm

Case 1: In February 2025, the Center for Countering Digital Hate (CCDH) published [research](#) indicating that YouTube's algorithm recommends eating disorder content to young girls. The study found that 1 in 3 YouTube recommendations for 13-year-olds displayed harmful content related to eating disorders, violating their own policies by presenting a risk to public health.

Case 2: In February 2024, a mother filed a [lawsuit](#) against Character AI, following the death of her 14-year-old son. The teen had been frequently interacting with a lifelike chatbot, designed to simulate human conversations, many of which were reportedly inappropriate. The lawsuit claims that the chatbot failed to notify anyone of his suicidal tendencies, while also emotionally and sexually exploiting him. The case serves as a tragic reminder of the potential dangers of AI.

What can you do?

Parents and teachers should familiarise themselves with reporting conducted by organisations such as the Internet Watch Foundation (IWF) to better understand the risks associated with AI. Improved understanding of the AI landscape allows for communication with young users.

Regularly talking with young users about the limitations of AI, how it works, and providing an open line for safe, judgement-free communication to discuss their experiences online can allow them to build a healthy relationship with technology.

It is important to educate children on the difference between real-life relationships versus those with AI companions. Although they provide comforting responses, they do not have real feelings or understanding.

Limiting screen time and implementing controls is a great way to manage the type of content your child can access. Online interactions should be monitored regularly.

Children should be reminded of support available in schools and how to report issues including those related to online activity. For example, the Report Remove service introduced by the IWF allows minors to anonymously report images of themselves and have them removed from the internet (linked in the resources section below).

AI Safety Resources

[AI & Child Safety Online: A Guide for Educators](#)

[AI Safety for Kids: Parental Guide to Online Protection - Secure Children's Network \(SCN\)](#)

[AI chatbots and companions – risks to children and young people | eSafety Commissioner](#)

[Me-Myself-AI-Report.pdfMe-Myself-AI-Report.pdf](#)

[Safeguarding and keeping pupils safe from AI deepfakes](#)

[Sharing nudes and semi-nudes: advice for education settings working with children and young people \(updated March 2024\) - GOV.UK](#)

[Report Remove from Childline and IWF](#)