Quick Look
AI and Online Safeguarding

Disclaimer: This resource is provided for informational and educational purposes only and does not constitute legal advice. If you require legal advice or advice on GDPR compliance, you should contact a suitably qualified professional. You should not rely on the material included within this resource and Twinkl does not accept any responsibility if you do.

Recent advances in technology mean that staff and children now have easy access to tools such as ChatGPT and Google Bard to produce their own content generated by artificial intelligence (AI). This is referred to as 'generative AI' and it creates significant online safeguarding challenges for those in the education sector. With just 62% of UK teachers reporting that they have never used generative AI, there is a compelling need to rapidly update online safeguarding practices and policies in schools.

This Quick Look provides a summary overview of the challenges presenting those with safeguarding responsibilities in order to help them understand the associated risks of using generative AI among staff and children.

1. What are the risks that generative AI poses?

- It stores and remembers any data it is given regardless of whether it is right, wrong, slanderous or inappropriate.
- It can create highly plausible and believable content that could give the impression that it comes from an authoritative or genuine source, e.g. demands for money, spam emails or highly offensive imagery or media.
- It can create a range of media that could be out of context or used without permission.
- Content referenced in generative AI can be out of date and/or unreliable.

2. What are the current safeguarding, online and GDPR factors that educators need to consider when planning how to integrate the use of AI into their practice?

- All existing Data Protection Act and GDPR protection principles must stay the same across any generative AI content created.
- Where generative AI content is used with children in school, it is the adult leading the activity who remains responsible for checking and quality assuring its accuracy and appropriateness at all times.
- Great caution is needed when using AI as a substitute for human therapists, educational psychologists or clinical judgements. There is a risk that individuals can share highly sensitive information with a machine that is stored in a dataset owned by an undisclosed entity. Additionally, the level of sophistication in AI chatbots may not be as advanced as perceived. This underscores the potential risk for AI to inadvertently worsen mental health concerns if auto-generated content is used to guide professional practice.
- Image creation is accessible to almost anyone on any computer. This poses a significant safeguarding concern, particularly in cases involving the production and dissemination of extremist ideologies or pornographic imagery.
- Evidence is emerging that AI is being used by young people when seeking advice on sensitive topics. Children will value the anonymity of AI 'advice', but be less able to discern its relative quality given their inexperience. This can lead to vulnerable young people following high-risk, auto-generated guidance. There is the danger that serious safeguarding issues could be missed by an adult where a child has used machine advice rather than speaking to a grown-up.

3. How can AI be used to reduce workload?

- It can generate potential starting points, ideas and suggestions for a lesson plan or unit of work.
- It can reduce time-consuming tasks, such as creating a worksheet of maths word problems linked to a history topic you might be teaching.
- Administrative tasks, such as a trip letter or an email, can be rapidly created for you to then edit accordingly.

4. How can educators use AI technology safely whilst being GDPR compliant?

- Any data entered into an AI system is stored. Don't enter data that can be identifiable. This covers things like names, contact details, emails and addresses, not just of children or parents, but of your school and its staff too.
- Be mindful of what information you're sharing at all times.
- Check everything that an AI system creates for you before use. While most platforms have some safeguards, human evaluation and judgement is needed of the completed items at all times. This includes things like auto-generated AI recommendations, advice, opinions and the factual accuracy of any content created.

5. What is the government doing to support education as schools adopt more digital tech into their practice?

- A Frontier AI Taskforce has been convened to fully evaluate the nature and scope of risk posed.
- Keeping children safe in education provides schools and colleges with information about what they need to do to protect pupils and students online and how they can limit children's exposure to risks from the school's or college's IT system.
- Schools are advised to refer to the DfE's filtering and monitoring standards to make sure they have the appropriate systems in place.

Links to further reading

Generative artificial intelligence (AI) in education - GOV.UK

Principles for the security of machine learning - NCSC.GOV.UK

The impact of the General Data Protection Regulation (GDPR) on artificial intelligence - European Parliamentary Research Service

Disclaimer: This resource is provided for informational purposes only. It is your responsibility to ensure that the information this resource contains is appropriate to use in your situation. This resource contains links to external websites. Please be aware that the inclusion of any link in this resource should not be taken as an endorsement of any kind by Twinkl of the linked website, or any association with its operators. You should also be aware that we have no control over the availability of the linked pages. If the link is not working, please let us know by contacting TwinklCares and we will try to fix it although we can assume no responsibility if this is the case. We are not responsible for the content of external sites.