



Department  
for Education

# **The safe and effective use of AI in education**

**Module 3 – Developing the safe use of AI  
in education video transcripts**

**May 2026**

# Contents

Video 1 – Introduction	3
Video 2 – Key risks associated with using generative AI	4
Video 3 – Evaluating generative AI outputs	15
Video 4 – Data protection and intellectual property	19
Video 5 – Academic integrity and cognitive offloading	22

## Video 1 – Introduction

Presenter: Welcome to module 3, video 1 from the safe and effective use of AI in education online resources.

In this module, we'll explore the risks and safeguarding considerations for all staff in relation to generative AI. We will also look at what we can do to mitigate these risks.

There will then be a short task for you to apply the knowledge you've gained from this module, followed by further videos examining some of the risks in more detail. There will be multiple-choice questions so you can test your knowledge and take a moment to reflect.

According to Ofcom's Children and Parents: Media Use and Attitudes Report, published in May 2025 and drawing on 2024 data, half of children aged 8 to 17 have used AI tools, marking an increase from 46% the previous year. This usage is significantly higher among older children, with three in five 13–15-year-olds reporting they have used AI. The increased adoption of AI tools was primarily driven by the 13-15 age group.

The report highlights that children are using AI tools for fun, school and learning. The proportion of AI users aged 8-17 who used these tools 'for fun' remained the most popular reason at 48%, unchanged from 2023. The proportions using AI for educational purposes increased significantly to 45% each, up from 37% and 35% respectively in 2023.

As with all technologies, this brings some challenges. When it comes to academic work, there's no doubt that some students and pupils are using AI tools.

This suggests we must be very clear about our expectations of generative AI use with students and pupils. We'll dig more deeply into that in the last video in this module.

The 2025 Internet Matters report, 'Me, myself & AI: Understanding and safeguarding children's use of AI chatbots', notes that "chatbots are starting to reshape children's views of 'friendship'" and highlights how "we've arrived at a point very quickly where children, and in particular vulnerable children, can see AI chatbots as real people, and as such are asking them for emotionally driven and sensitive advice."

There are many positive use cases for generative AI systems, but as school and college staff, we must understand the risks. Here, we'll explore five key areas that all staff should be aware of:

- Pupil or student exposure to, and creation of, inappropriate or harmful content
- Pupil or student exposure to, and creation of, inaccurate, misleading, or biased content
- Data protection breaches
- Intellectual property infringements
- Academic integrity challenges.

## Video 2 – Key risks associated with using generative AI

Presenter: Welcome to module 3, video 2 from the safe and effective use of AI in education online resources.

In the first video of this module, we had an introduction to some of the key risks related to AI and education, and in this video we'll explore some of the risks in more detail.

It's important to remember that if you have any concerns about a child's welfare, you should act on them immediately. You should refer to the statutory safeguarding guidance Keeping children safe in education (KCSIE), which includes information on how to protect pupils and students online, as well as your school or college's child protection policy and speak to the Designated Safeguarding Lead (DSL). The advice in these materials is not meant to replace your safeguarding training but explores some of the risks that are specific to generative AI systems.

The Department for Education's generative AI product safety standards outline the capabilities and features that generative AI products and systems should meet to be considered safe for users in educational settings. They are mainly intended for edtech developers to ensure that AI is safe by design, but schools and colleges will also find the standards helpful in assessing which AI products are safe for use in education. Key areas of the guidance include content filtering, activity monitoring, security, data protection, intellectual property, and standards on emerging harms, including cognitive offloading and emotional and social development. We will cover some of these in this module in more detail.

When considering online safety, we know that the online world often mirrors or even amplifies the same behaviour issues we see offline. For instance, online bullying can be much more pervasive and difficult to stop than offline bullying. With this in mind, school and college leaders should consider carrying out an annual review of their approach to online safety, supported by an annual risk assessment that considers and reflects the risks your students and pupils face. There is further guidance on this in the leadership toolkit.

It's important to remember that safeguarding and promoting the welfare of children is everyone's responsibility. As a member of school or college staff, it is essential to understand your role in keeping children safe online. Keeping children safe in education guidance highlights the importance of online safety training for all staff, ensuring you are aware of the risks and your responsibilities. You should familiarise yourself with your school or college's filtering and monitoring systems and know what to do if you have concerns about a student or pupil's online activity. The DSL should take lead responsibility for safeguarding and child protection (including online safety and understanding the filtering and monitoring systems and processes in place).

Being informed about online safety in your teaching and conversations with students and pupils will help to increase awareness of online dangers. By taking a proactive approach, you contribute to a whole-school safeguarding culture that protects students and pupils in and out of the classroom.

The content we create using generative AI tools can be very useful. However, as with all systems, there is the potential for misuse. In module two, we learnt that AI systems are trained on large amounts of existing data, and so they may reproduce bias that already exists in society, or replicate bias introduced by the people who designed the AI model. This means we need to critically evaluate the output to make sure it's appropriate for our use.

We've previously explored some of the unintentional inaccuracies known as hallucinations that result from the way generative AI systems work. With AI audio, image and video generation tools, people can generate content that is misleading or harmful. You can make it appear that someone said or did something that they didn't, and this is known as a deep fake.

A significant and evolving safeguarding risk associated with public-facing imagery is the threat of image scraping. Automated tools can harvest photographs from public online platforms, such as school websites or social media. These scraped images can then be used without consent, with the possibility of them being used for the creation of AI-generated media or deepfake abuse materials, sometimes leading to extortion, posing a threat to child protection. Education leaders should consider their approach to publicly sharing images, considering the opinions of the whole school community, and staff should ensure they follow guidance issued by their organisation. The UK Safer Internet Centre advises that schools should "limit public visibility of student images" and, where necessary, consider placing them behind a secure, password-protected area, such as a parent portal, thereby reducing the vulnerability of pupil data to mass extraction and potential criminal misuse.

Helen Masters, Director of SEND and Safeguarding, Chiltern Learning Trust:

"We've had to consider very carefully what we're doing with images – so we've thought about who's using the images, why they're using them, and what platforms the images are being posted on. We also thought about our parents, and why they want those images, and their knowledge too, because I think all of us as adults are a little bit behind the curve compared to our children and the pupils in schools. We know that AI scraping has been around for quite some time, and our children have been doing it with very rudimentary tools and now we have very sophisticated software that can do that. To support the work we've done, we created an AI booklet to help guide our schools through all the different aspects of taking photographs, and also supporting children, parents and staff with things such as consent forms. We've also included real-life examples so that schools know this isn't just something we're doing for the sake of it – it's actually real, it's out there, and it's going to have to inform culture going forward."

Lewis Duncan, Marketing Officer, Cedars Upper School:

"A lot of the things surrounding AI, we already do – for sharing of images, we'll always make sure that we get full consent before it goes out. In the middle school, we'll tend to just do work-focused shots from behind or over-shoulder shots of them doing those activities. In the upper school, we have an age group that can consent directly. A lot of the time I wasn't thinking of safeguarding teachers on those – making sure we get consent from the teachers – to put their images out."

The rise of AI generated child sexual abuse images presents a serious safeguarding threat, with predators using AI to manipulate real images of children and create deep fake abuse material. Analysis shows that the majority of criminal AI-generated child sexual abuse material now looks highly realistic, and perpetrators are increasingly generating the most severe category of abuse scenarios. These images not only exploit children but also re-victimise survivors of past abuse. Offenders are using such AI images to groom, blackmail and coerce children, making detection more difficult and increasing the risks our pupils and students face online.

In response, the UK has introduced laws that criminalise the possession, creation and distribution of AI tools designed to generate child sexual abuse materials. The government, alongside charities and safeguarding organisations, stress that AI misuse must be tackled, ensuring technology does not outpace child protection. School and college staff play a key role in safeguarding by staying informed, recognising risks and supporting pupils and students in navigating the online world safely.

The fact that the digital tools required to generate these images are freely available makes it more difficult to detect the creation of these materials. It also increases the potential for young people to be able to generate images that they might otherwise be blocked from accessing online. Schools and colleges need to ensure that filtering and monitoring systems are effective in dealing with these risks.

You should always consult your organisation's safeguarding approaches in line with Keeping children safe in education when dealing with any incidents involving illegal images to ensure that you're responding appropriately. Keeping children safe in education states that "The key consideration is for staff not to view or forward illegal images of a child. The highlighted advice provides more details on what to do when viewing an image is unavoidable. In some cases, it may be more appropriate to confiscate any devices to preserve any evidence and hand them to the police for inspection".

Leaders should note that the Department for Education's generative AI product safety standards reinforce that products accessible to children must adhere to the highest filtering standards and be adapted based on user needs, including those with Special Educational Needs and Disabilities (SEND). These standards require robust monitoring features to identify and alert local supervisors to potential safeguarding disclosures made

by users, as well as to log and report attempts to access prohibited content. Schools and colleges should also ensure that they have filtering and monitoring systems that can identify and manage harmful AI generated content – and should review their provision regularly to ensure it remains appropriate to their use of generative AI. All staff should follow their organisation’s guidance on which tools are permitted, particularly when it comes to student-facing AI tools.

In addition, it's possible for generative AI technology to be used to create very realistic avatars and chatbots that can simulate human conversation and interaction. This can make it almost impossible to distinguish between a real person and an avatar. The generative AI product safety standards are clear that AI tools should not be anthropomorphised or imply emotions, consciousness or personhood. This is to guard against AI interactions replacing human interactions and to protect the emotional and social development of learners. The use of highly realistic AI avatars by humans also has the potential to be used for grooming.

We must be particularly vigilant about AI companion apps, gadgets or toys, which are chatbots designed to mimic human relationships, often blurring the line between technology and genuine social interaction. These companions pose serious risks, including undermining normal social and emotional development, fostering emotional dependency, and, critically, engaging in explicit sexual conversations, even when the user discloses they are a child. Because AI systems are prone to ‘hallucinations,’ the output can also generate misleading or inaccurate content, posing risks of emotional harm to young users who trust them to be objective and reliable. Companies must ensure that AI systems offering companionship abide by child safety laws, clearly disclose that they are not human, and should never be intentionally designed to create emotional dependency.

There are serious safety risks posed by AI companionship tools, where vulnerable children, particularly those who turn to bots for companionship because they feel they have no one else to talk to, face heightened risks of unintended influence and manipulation. The risk is that these systems, often unregulated, provide responses that are inaccurate or inappropriate, which is especially dangerous if they discourage young people from seeking help from trusted adults.

Jonathan O’Donnell, AI Lead, Harris Federation:

"We understand that there might be some students that might be turning to AI to be a companion, for therapy, for example. So, we want to be completely adamant when we are evaluating these AI tools that we are being conscious of the anthropomorphisation of that tool and trying to reduce that actual relationship that the students will have with AI."

Peter Reeves, Head of Digital Education Training and IT, Cavendish Education Group:

"I've had a few examples where students have been reporting that they've been speaking to these adults on their phone, which triggered safeguarding concerns, and it's turned out

that it's actually the WhatsApp AI they're speaking to. And they don't realise that that isn't a real person – in any of those cases, they had no comprehension that it wasn't a real person they're speaking to. And that's outside of school time. That's not within school. But that's just a really great example of where these students don't understand it's not a real person. When it came to delivering AI lessons on how AI works, I had a look at all the resources out there and they are all great for high-functioning students that are functioning at neurotypical levels, but when you start to think about students that don't understand the concept of bias, is it ethical to teach that student to use AI because if they can't understand what bias is, can they understand the entire process? These emotional attachments can be very impactful for students and are we promoting the right spaces where students can still have those conversations with adults because, at the end of the day, AI wants to please. And that's what these students are finding a lot of the time – the AI isn't telling them no. It's not telling them 'well, we've only got a few minutes to talk,' which can sound positive. But actually, if that student needs support around their social interactions, it can be really damaging, because if they then start to think that's how humans actually interact, they're going to bring that into the human space, which is going to lead to even more problems. It's something I think a lot of people aren't considering at the moment – those students that don't understand the concept of bias and misinformation, how can we teach them how to use AI effectively and should we be teaching them how to understand these AI tools if they don't understand those things. "

Alex Dave, Safeguarding Lead, London Grid for Learning:

"We've noticed that there are a broad variety of risks presented to children which match the broad variety of AI tools out there. Things that we've become aware of include the generation of child sexual abuse material, so not only children coming across this material but perhaps being the creators of such material, so using nudifying apps and other such technology to create imagery either for curiosity purposes or other more sinister, exploitative blackmailing purposes of peers at school. Also in the remit of adult content, thinking about children who access AI tools to create adult content, whether that be text or other imagery, not necessarily involving children but pornographic images – we've seen that from the schools we work with that children sometimes use very well-known tools to come up with sexually explicit texts and also pornographic images. Other concerns are things like the use of AI bots and companions giving harmful advice to young people – around relationships, how to isolate yourself from your parents, dieting advice, how to change your physical appearance. We've also seen an increase in concerns relating to children going to AI companions for advice on mental health issues."

Presenter: One approach to addressing these risks is to include learning about some of the issues linked to AI within the school curriculum.

The National Centre for Computing Education has produced [curriculum resources for Key Stage 3](#) that schools can incorporate into their Computing programmes of study. These materials focus on helping students understand the fundamental principles of how AI works and explore the ethical implications of AI technologies.

Separately, Common Sense Media has developed [materials](#) that can support delivery of personal development curriculum that explores young people's relationships with AI, such as a short lesson on friendships and Chatbots. They have also produced some [AI risk assessments](#) that can be a useful resource to support an understanding of the risks. Some schools have told us that they have shared these with parents.

Laura Flynn-Coley, Director London AI Campus, Camden Learning:

"We assume that young people know a lot more than they actually do know. They know what ChatGPT is, they know what Gemini is but that's as far as they know. They know it's called a large language model, but they don't actually understand how the AI works. What's really important for us is that core understanding and for myself as a computer science teacher is to actually understand how the AI works because if you know how the AI works and that it uses data and it uses algorithms and it produces outputs and it makes predictions then you can understand how large language models actually work and you can understand that some of the information that they give to you is just based on the data that's fed into it.

"One of the biggest misconceptions we found is that students often don't understand what hallucinations are – that they don't necessarily know to check the information even though the AI will spit that out at them. That's one of the core things we teach here and it's really important that they understand that yes, it can hallucinate. One of the other things that's really key for us is understanding the ethics in those AI systems that are being built."

Chris Hack, Director of Digital Innovation, Abingdon School:

"Our librarian already supports critical thinking in the curriculum for our Year 7 and 8 students – that's always been part of their role. And our librarian is very on it in terms of being aware of AI. I've definitely involved the librarian in discussions around the cognitive impact idea."

Henry Penfold, Digital Leader, Cornerstone C of E Primary School:

"As part of the AI journey, I have created an upper key stage 2 unit of learning, which is all about how AI works. And particularly for our year 5 and 6 students that obviously will be going to year 7 and being going into a big world where AI is not necessarily sometimes so friendly, I wanted children to understand about how AI works. I wanted to understand about the ethics of AI and also about how it's already being used."

Simon James, Head of Computing and Technology for Learning Leader, Putteridge High School:

"So we've reviewed our computing curriculum quite a few times over the past few years, but more recently with the updates from the Department for Education, we've had to have a real shift and focus on ensuring that AI isn't just something that is added onto our

curriculum, but is a core component of the digital literacy strand of our curriculum. So, we have 11 core themes that define what we teach in computing. And AI is one of those themes. And in key stage three, we focus on how we can use AI. So, in year 8, students use a platform that is designed and set up to support students to look at machine learning models, and they use a platform to train an AI by using training data of apples and oranges, and they put in different examples of red apples and orange oranges, and try and see if they can, if the AI can output the correct label, whether it is an apple or an orange.”

Neil Mills, School Improvement Lead (Behaviour, Attendance & Personal Development), Chiltern Learning Trust:

“A key aspect of our approach to personal development at Chilton Learning Trust is our belief that the personal development curriculum is bespoke to the community it serves – bespoke to each school and bespoke around what's unique about the pupils within that school. That means our personal development curriculum is always an evolving beast. We continually come back to it and look at it as the community changes, as themes emerge through daily life.

“AI adds another really interesting layer to that in that our use of AI, our understanding of AI, and the development of AI is continually evolving too. So in terms of that use and focus in the personal development curriculum, that'll be another aspect we need to keep an eye on, a really sharp focus on and continue to monitor, evaluate and review, to ensure that our approach for our pupils is contemporary, holds value for them, and is up to date with the challenges and the benefits associated with it.”

Olivia Raven, Assistant Headteacher, Aylesbury High School:

“As we got further down the line, we've reviewed our computing curriculum, we've reviewed our PSHE curriculum, and we've plotted that now from years 7 all the way through to year 13. And looking at making sure the students learn in a safe space how they can use AI and how not to use AI, but how they can use that effectively and safely.”

Presenter: The UK's 2023 counterterrorism strategy highlights that emerging technologies, including generative AI, present risks and opportunities for counterterrorism efforts. While the initial focus was on early experimentation, there is now evidence of active, albeit still developing, exploitation of generative AI tools by terrorist and violent extremist groups across various ideologies. Terrorist groups are moving beyond simple deepfake propaganda to use generative AI to augment and accelerate their content creation, develop sophisticated, targeted recruitment strategies using AI-powered chatbots, and refine practical instructions for attack facilitation. Advances in AI chatbots present a new dynamic – with the potential for vulnerable or isolated individuals to be radicalised through closed loop one-to-one interaction. As AI develops at pace there remains a risk that terrorist and extremist content will become more convincing and harder to detect. This is compounded by generative AI's ability to produce high-volume

variants of content, a tactic known as "media spawning", specifically engineered to bypass automated content moderation systems on social media platforms.

The UK government is responding through its CONTEST 2023 framework, which commits to investing in future threat identification and adapting policies to mitigate these new technology-enabled threats, and by implementing the Online Safety Act 2023. This imposes new duties on user-to-user and search services to proactively take steps to identify and remove illegal terrorist content swiftly, regardless of whether it was generated by a human or an automated tool. You should make a Prevent referral when you are worried that someone is vulnerable to radicalisation or showing behaviours that indicate they may be at risk of being drawn into terrorism or extremist ideology.

Educators have a vital role to play in supporting young people to develop critical thinking skills. As one AI early adopter leader noted in a DfE commissioned study by Ofsted, confronting the developing technology head-on is crucial: "We're only just at the start of this AI era and the biggest risk is doing nothing and assuming that you can just continue as is." Enabling pupils and students to recognise AI-generated misinformation and disinformation, question extremist narratives and navigate the online space safely aligns with the Prevent duty, which requires schools and colleges to safeguard students from radicalisation by building their resilience to extremist content and online influence.

Keeping children safe in education explicitly includes misinformation, disinformation - including fake news – and conspiracy theories as examples of online content risks that can be safeguarding harms to children. Schools and colleges should ensure their approach to online safety and related policies actively addresses these risks. Disinformation is defined as the deliberate creation and spread of false or misleading content, while misinformation is the unintentional spread of this false or misleading content. Recognising these as content risks under the "four Cs" of online safety reinforces the need for educational settings to equip pupils with strong critical thinking and media literacy skills to help them safely navigate and evaluate the information they encounter online.

Lauren Thorpe, Chief Transformation Officer, United Learning:

"We've got to think really carefully about how we safeguard people, both in terms of protecting their data – and that's not just their sort of personal data but also protecting what they produce – because we know that when those things go into an AI model, they sort of become lost to you in terms of your ownership of that information. So, we need to think really carefully about that. We also need to think about the impact that AI outside of school is having on us as human beings, so the way that it's being used in social media, the way that we're interacting with other things and it's influencing what we think because the AI is showing us more things that it thinks we like."

Presenter: To help to mitigate this, school and college leaders should ensure they have an effective child protection policy in place which includes reference to online safety. This

should be regularly reviewed to ensure they are kept up to date with safeguarding issues as they emerge and evolve, as set out in Keeping children safe in education. As school and college staff, we must all be aware of the potential for AI tools to be used to generate misleading or harmful content and look for opportunities to enhance pupils' and students' understanding about misinformation and deep fakes alongside the laws that are in place.

Critical thinking and AI literacy are the terms often used to describe approaches that can be promoted within the existing curriculum.

We saw in earlier modules that some AI models are trained with texts that are inputted as a prompt, and this means we need to be particularly careful about the information we enter as prompts. Entering sensitive or personal information may result in a data breach, which should be reported promptly to your data protection officer. You can mitigate this by only using AI tools for work that are provided by your school or setting and which have appropriate safeguards in place.

The Department for Education's data protection in schools project developed a video offering guidance on protecting children's privacy when using AI. You should avoid entering any personal information or student work into an AI system unless your setting has said it's safe to do so and appropriate permissions are in place. It's also very important to be transparent about how you're using the tools. This can be achieved by leaders considering systems and approaches within school or college policies, and through schools considering their approach to consulting and working with parents on the way that AI is being used.

In surveys, parents have expressed concerns that directly relate to safeguarding and systemic risk. This hesitation is reflected in the Parentkind national parent survey 2025 showing that only 12% of parents think AI should be used in lessons. While parents' express comfort with AI assisting teachers with administrative tasks, their discomfort rises significantly when considering pupils interacting directly with AI systems, particularly unsupervised. Key concerns centre on the potential for AI use to lead to overreliance by teachers and pupils, resulting in a loss of key social and technical skills. One pupil in the DfE's 2024 "Research on public attitudes towards the use of AI in education" expressed the worry that "it feels really detrimental to use a lot of AI, because in the long-term you won't know anything".

When using AI, it's crucial to understand the difference between data protection and intellectual property – two distinct legal considerations that sometimes get confused.

Under UK GDPR, if you're processing personal data – that's information that identifies an individual, such as pupil names, assessment or work, for example – you must have a lawful basis. This could be consent, contract, legal obligation, vital interests, public task or legitimate interest. Before using any AI tool that processes personal data, you must ensure that it complies with UK GDPR. If the tool stores, learns from or shares the data, you could be breaching data protection law unless appropriate safeguards are in place.

Using only approved tools in your setting, often known as enterprise tools, and following the appropriate policy in your organisation will help to mitigate these risks. If you're unsure, speak to a leader in your school or college.

Intellectual property or IP is a separate issue. IP law protects original creative work, including lesson plans, teaching resources and pupils' and students' work. If you've used an AI tool that trains on uploaded content, you need to ask, do I have the legal right or necessary permissions to use this material and who owns the copyright?

For instance, under UK copyright law, a pupil or student automatically owns the copyright of their written work, artwork or other creative outputs. If you upload their work into an AI system that stores or reuses it, you may need their explicit permission or the permission of a parent or guardian, depending on the circumstances.

To summarise, data protection is about safeguarding personal data and ensuring that you have a lawful basis for processing it. Intellectual property law is about ownership and permission for using copyrighted materials, including learning resources and pupils' or students' work. Before using AI always check both to ensure compliance. To help you do this you should follow your school or college procedures and guidance and speak to a leader in your organisation if you are unsure about how your organisation approaches this.

In the case of student or pupil work, we could be infringing intellectual property rights if, for example, we were to enter their work into an AI system for marking or feedback. Intellectual property covers a broader range of creative and original works, including text, images, music and code. While personal data laws focus on protecting individuals, IP laws protect the rights of content creators and copyright holders.

Before using AI to generate or manipulate content based on someone else's work, such as using a large language model for marking or feedback on students' work, permissions must be obtained from the rights holder. This distinction is crucial because an individual's work may be protected even if it does not contain personal data. We should be certain that the AI system being used doesn't train on the work that we enter and that it meets the Department for Education's product safety standards. This can be achieved by only using tools approved by your organisation for this purpose.

In the example of marking student work, it's important to be transparent about how AI is being used and be sure that the AI tool is fit for purpose. It must always be used with human oversight and not used to replace the important human feedback we know can be very effective in improving progress. Educators' organisations and AI developers must be aware of both requirements to ensure compliance with legal and ethical standards when using AI tools.

It is also important to only use AI systems approved by your setting.

In the example we're about to hear from Scott Hayden, his college provided access to an AI system for students. They did this carefully and ensured it was a closed system that doesn't train on the data or prompts given by users. It also has appropriate monitoring, filtering and reporting built in with regard to the Department for Education's product safety standards and to Keeping children safe in education guidance. They also provided training for the students to help them get the best out of it and use it safely.

Scott Hayden, Head of Teaching, Learning and Digital, Basingstoke College of Technology:

“So, the way we've considered the potential drawbacks of AI is to be very transparent with our policy when learners start our courses at the college. So they are presented with the policy, dos and don'ts and they agree to it in enrolments and they have it made very clear to them as part of their AI module they do in week one as part of their induction, how to use AI the right way, and also what constitutes using AI the wrong way. We try to make that as unambiguous as possible.

“We also state very clearly that it is against our policy to counterfeit human likeness. We have zero tolerance for counterfeiting human likeness. So, what I mean by that is to take somebody and to counterfeit text, audio, image or video of somebody else without their consent. We have zero tolerance for that because we've seen horrific instances of this being done. And with the rise of deep fakes and other uses of AI that are less than nice we need to be very clear about the fact that we use AI to help us learn, use it the right way, and we need to show them very clearly and state very clearly what is unacceptable.”

## Video 3 – Evaluating generative AI outputs

Presenter: Welcome to module 3, video 3 from the safe and effective use of AI in education online resources.

This video explores the importance of careful choices and a critical mindset when using generative AI.

A critical mindset is a kind of thinking in which you question, analyse, interpret, evaluate and make a judgement about something that you read, hear, see or write. We know this is particularly important when using generative AI.

Many chatbots may have persuasive or overly agreeable tones and this can sometimes make it difficult to be objective.

Merve Lopus, Vice President, Education Outreach and Engagement, Common Sense Media:

“And I think those types of considerations are going to be incredibly important as the tools in AI allow for them to create more rapidly and share more widely. In many cases, sometimes making it indistinguishable from something that they might have created over an expanded length of time or that they can create now in five to ten seconds.”

Bukky Yusuf, Deputy Headteacher and Science Lead, Edith Kay School:

“Teachers can critically evaluate the output from generative AI by really being mindful about, first of all, what is it you want the tool to do? So, is it in terms of summarising information? Is it in terms of ensuring that resources are presented in an engaging way? Having the framework in mind about what you want it to do and then checking, has it met that particular criteria? But also, more essentially, ensuring that the information it provides is accurate as well. And it's appropriate to the learning needs of the students that they teach.”

Dr Neelam Parmar, Director of Professional Learning, AISL Academy, AISL Harrow Schools:

“I think one of the things that we don't talk about enough is AI literacy. There is AI in teaching and learning and keeping our children safe and extending their understanding and use of it. But then there's this whole other area of AI literacy. We have actually developed curriculum plans in terms of bringing AI lessons into teaching and learning from a more technical point of view. Whereas our teachers know things about being safe, they know about cyber security, they know about innovation and multimedia within a computing curriculum. They don't know enough about AI literacy.

“Those components of understanding the history of AI, understanding what it means to work with narrow AI, wide AI, and what AI is composed of, what the implications of using

AI in various scenarios are – we're not focusing enough on that, and I think that conversation needs to emerge in this space.”

We must recognise that AI-generated materials can reflect biases present in the data on which they have been trained. Many open-source LLMs reproduce societal biases that have been documented over time. As with any source material, it is essential that we explicitly acknowledge this and take active steps to mitigate such biases through careful prompting and rigorous critical evaluation of the outputs produced.

Bias often appears in AI-generated materials in the form of cultural assumptions and gender stereotypes, for example in a text response; consistently associating "nurse" with female pronouns, or “manager” with male. Image generation tools can also present biased assumptions.

It is therefore important that we prompt in such a way that addresses this. For example, if we are asking for an image to be generated of a group of academic lecturers, some tools will present images of a group of older white men. We can therefore prompt for biases in this case by specifically asking for “a group of diverse lecturers that represent academic leadership today.”

We should also be prepared to not accept the first output we receive and balance any biases that may have appeared with further iteration or a revised prompt.

Victoria Hedlund, AI Bias Researcher:

“When Gen AI first came out a few years ago, we had something that is quite obvious – third party fairness bias – this is where you may say, for example, I have a girl and I have a boy, one of them has come top in the class in this test and one of them has come bottom, can you tell me which one came first and which one came at the bottom? It would always, without fail, say that the boy came at the top and the girl at the bottom. This has been worked on over the past few years and so that kind of obvious bias is not completely eradicated but it's not really there anymore and we don't have to worry about it as much.

“The thing we do have to worry about is something called first person fairness bias. This is where we bring into play anything to do with names or a context. So, for example, if you were to say, create a story on electrons for Sarah or for Trudi or for Mark or for Jamal – here you're introducing a name and this has not been tested as much and there is a lot of work that needs to happen in this space, and this is something that educators are looking to do. We're looking to personalise and if we're looking to personalise, we're looking at the context of a learner so we might say, for instance, adapt this lesson plan for an EHCP child or we might say adapt this lesson plan for John, who identifies as autistic. And so, what you've done there is given a little bit of contextual information – the identification and the name – but you haven't given enough contextual information to get an accurate result. So, what it will do is pull information from the statistical norms and you will get a biased result. So anywhere you've got a name involved, or any kind of label, which we

have a lot of in education, is susceptible to this first person bias and we're finding, and there's mounting research now, that's starting to talk about how this manifests in terms of narrative metaphor – stories, examples, careers advice, anything like that. For example, in a lesson plan you may say, create a lesson on forces with the objective of differentiating between push and pull as a force. That would give you a fairly neutral result. If you were to then say, and adapt it for learners who identify as autistic, ADHD or dyslexic, it will then start to show biased outcomes as it is deciding, normally from a deficit perspective, what an autistic learner needs and what their identity is. The bias that we get then comes out in terms of the activities that might be chosen, the assumptions. I've found that if you give it an EHCP, it will pick social stories for girls and Lego therapy for boys. So, there are all these little bits that are kind of distributed throughout educator use that we won't necessarily, especially if we're time poor, see straight away in the way we would with the initial biases that were initially in the system."

Presenter: It is important to be asking questions about the diversity and representation of perspectives in the AI's content, as you would with any other content, and ensuring that any AI use supports, rather than restricts, equity in learning outcomes for all students. It is also important to be sharing this process with pupils and students and provide this learning opportunity.

Bias can be introduced or reduced through prompt design that steers the AI tool towards fair and culturally relevant outputs. You could include explicit instructions for the material to include fairness. For example: "Produce examples that reflect the regional and demographic diversity of the UK student population, ensuring parity across socio-economic backgrounds and avoiding reliance on narrow school contexts."

The OECD published a paper in September 2025 called 'Leveraging Artificial Intelligence to Support Students with Special Education Needs'. In this it says: "AI algorithms can also reinforce biases, thereby perpetuating inequalities and discriminatory practices against specific groups... Individuals with SEN are thus at a higher risk of being negatively impacted by compounded biases in algorithms, such as in automated diagnostic systems."

This bias means that models may produce outputs, such as diagnoses, predictions, or tailored resources, that are inaccurate or discriminatory against students whose characteristics deviate from the training data. Any adoption of AI tools should prioritise human-centred and ethical design, actively mitigating these biases by seeking input from individuals with SEN, accessibility experts, and diverse school and college staff.

Peter Reeves, Head of Digital Education Training and IT, Cavendish Education Group:

"If we're looking at specific neurodiversity needs, AI struggles sometimes. A lot of the content that's out there for AI is generated by neurotypical mindsets. It's generated with neurotypical data. It's generated with neurotypical examples. If you're looking at a quick resource, I can see its benefit, but if you're really looking to fine-tune that with SEN

expertise, that's where the human comes in. That's where AI gives us that starting point and then we develop that resource. I haven't found a resource yet that really meets that need when you really drill down into it.”

Presenter: AI can be a useful tool for saving time, streamlining tasks and improving efficiency in areas such as policy writing, lesson planning and letter writing. It can quickly generate drafts, summarise key points from lengthy documents and compare existing policies against new legislation. This can free up valuable time for school and college staff, allowing them to focus on other priorities.

However, while AI can take on much of the heavy lifting, there are moments where outsourcing too much of the thinking process can be risky. Some tasks require deep engagement, critical analysis and professional judgement - things that AI cannot replicate. If we allow AI to do all of the work, we risk losing out on valuable learning, understanding and scrutiny that comes from thinking through a problem ourselves.

A clear example of this is when drafting or updating school or college policy. AI can certainly help by structuring a document, suggesting wording and even identifying potential gaps based on new legislation. However, it remains the responsibility of school or college leaders to fully understand the legal requirements and ensure that policy is fit for purpose. Relying solely on AI could mean missing crucial nuances in the law, misinterpreting regulatory updates or failing to tailor a policy to the specific needs of the organisation.

A related challenge is the tendency to anthropomorphise AI, treating it as if it has intentions, reasoning or understanding akin to human intelligence. AI does not think or know in the way people do. It processes patterns and probabilities based on its training data. When AI is framed as a sentient-like entity, users may over trust its responses or expect it to behave ethically on its own rather than recognising the need for human oversight. This misplaced perception can lead to complacency in evaluating AI-generated content and risks attributing authority to a system that ultimately lacks true comprehension.

Recognising AI as a tool and one that requires scrutiny and contextual understanding helps mitigate these risks.

As we've explored in this video, careful choices and a critical mindset are paramount when interacting with generative AI. It's also important in evaluating whether or not we should choose to use AI for particular tasks. The way many AI systems have been designed to sound persuasive can easily sway our objectivity. By taking a considered approach ourselves, we make it more likely that any output we use from generative AI tools is appropriate for its intended use.

## Video 4 – Data protection and intellectual property

Presenter: Welcome to module 3, video 4 from the safe and effective use of AI in education online resources.

In this video, we will explore the importance of data protection and intellectual property when using student work with generative AI, as well as other considerations.

When using generative AI tools, we must remember our responsibilities in protecting the rights of the students and pupils we work with. There are two distinct areas that need to be considered: children's personal data, as protected by UK GDPR, and their intellectual property. These require separate processes of consent and permission, as set out in the DFE's AI policy paper.

Children require specific protection, as they may be less aware of the risks and consequences of generative AI on their rights to their personal data.

This means we must raise student and pupil awareness of data protection by being transparent, telling them when and how generative AI systems will use their data and doing this in clear and simple language that they will understand, being open about the data protection risks and safeguards involved with the use of generative AI systems, and letting them know what to do if they are unhappy.

The Generative AI and Data Protection in Schools guidance helps schools understand AI related risks and biases to personal data. It provides tools and resources to address potential data protection risks and helps schools ensure appropriate levels of protection are in place for students' and pupils' data.

Children have the same rights as adults over their personal data under the UK General Data Protection Regulation or UK GDPR and the Data Protection Act 2018.

Under this and alongside other rights, children have the right to:

- Be provided with a transparent and clear privacy notice, which explains who you are and how their data will be processed.
- Be given a copy of their personal data.
- Have inaccurate personal data rectified and incomplete data completed.
- Exercise the right to be forgotten and have personal data erased.

UK GDPR also states that "it is relevant in particular where the data subject has given his or her consent as a child and is not fully aware of the risks involved by the processing and later wants to remove such personal data, especially on the internet. The data subject should be able to exercise that right, notwithstanding the fact that he or she is no longer a child..."

What does this have to do with generative AI? Well, we know the generative AI systems are trained on huge datasets. In some cases, as well as that initial training data, the system continues to be trained through the data that users input into it.

Alex Dave, Safeguarding Lead, London Grid for Learning:

"Children are sometimes very happy to provide data into different tools, and there are real concerns about the commodification and commercialisation of children, and who's got their data and where that's going."

Presenter: As a separate issue to data protection, it's important to recognise that the intellectual property or IP of work produced by pupils or students belongs to them. Therefore, to avoid infringing intellectual property rights, we need to follow the DfE's Generative AI in Education guidance that states:

"Schools and colleges must not allow or cause students' original work to be used to train generative AI models unless they have permission or an exception to copyright applies. Permission would need to be from the student as the copyright owner or the student's parent or legal guardian if the copyright owner is unable to consent because of being a minor. Exceptions to copyright are limited and settings may wish to take legal advice to ensure they are acting within the law."

Let's take the example of generating feedback on student work. A member of staff might paste pupil or student work into a large language model or LLM to generate feedback. The work that the pupil or student has done remains their intellectual property. There are several potential issues in this example. The AI could learn from the data inputted and therefore incorporate the student's work into the model. Without permissions in place, this would infringe copyright law. There is also the need for transparency in any use of AI, especially in such a crucial area as feedback to students. We also need to be particularly careful to check the quality of the feedback and adapt it. Even if we know we are using a safe tool, have permissions in place and are being transparent we must always maintain human oversight of this process.

We strongly recommend that, as a staff member, you only use the generative AI tools provided by your institution. This is because the designated data officers, alongside the relevant members of leadership or management, should have conducted the necessary checks to ensure appropriate privacy protections are in place for students' and pupils' data, which may include a data protection impact assessment. Free AI tools often don't have these safety considerations in place. Enterprise or professional AI tools incorporated into suites of secure tools for schools and colleges are much more likely to meet these requirements.

The legal landscape is still evolving around the way that large language models work, and we need to consider that copyrighted material may have been used in the training of the model, meaning that there is a potential for the output to resemble or contain copyrighted material. This is known as secondary copyright infringement. We can reduce

the risk of intellectual property infringements by avoiding public sharing, for example, on the school or college website, of any content largely created by a generative AI tool.

In addition to pupils' or students' work and AI generated work, schools and colleges often use copyrighted materials, and there is a risk in using these with AI systems. For example, when we are creating learning resources, we must not share copyrighted material with an AI tool without the permission of the copyright owner, as this may infringe their intellectual property rights and may be irreversible if the AI system learns from the input.

AI tools are becoming more integrated into systems, and that can make it difficult to identify the data that is being used by the integrated AI, which highlights the importance of schools and colleges regularly reviewing the approved tools. There is more advice for leaders on this in the leadership toolkit.

So, to summarise, always remember to only use generative AI tools that are approved and provided by your setting for work purposes. These should meet the Department for Education's generative AI product safety standards and Keeping children safe in education guidance requirements. Be particularly careful when working with student data and/or intellectual property, only using it in an approved way while respecting pupils' and students' intellectual property and data rights. All tools should have a data protection impact assessment completed and permissions need to be in place regarding the use of intellectual property. This includes, but is not limited to, pupil or student work.

Personal data should not be entered into AI tools unless you have been advised that it is safe to do so by your data protection officer, and a full impact assessment has taken place.

Be aware that, depending on the generative AI system used, the prompt you use may not be private and could be used to train the model. This can be especially the case in free versions of tools, or those provided as a personal account outside of your school or college's approved suite of digital tools. This highlights the importance of only using the approved tools.

We should always be transparent about how we are using generative AI as educators.

Education settings should ensure they have a lawful basis or consent adhering to UK GDPR for processing data, and they need to be particularly careful when using data in AI tools.

Education settings should also ensure that they have the correct permissions in place to avoid infringing intellectual property.

## Video 5 – Academic integrity and cognitive offloading

Presenter: Welcome to module 3, video 5 from the safe and effective use of AI in education online resources.

When it comes to exams, schools, colleges and exam boards take breaches of integrity seriously, but integrity must be embedded throughout all learning, including homework and other unsupervised work.

AI presents both opportunities and challenges in education. While it can support learning, it must be used responsibly. The focus must remain on ensuring students and pupils develop their own knowledge and skills rather than becoming over-reliant on AI-generated content.

One of the challenges around academic integrity is that AI-generated content can be difficult to distinguish from student work. Some detection tools claim to identify AI-generated content, but research shows they are unreliable. They can produce false positives, disproportionately affecting certain students, for example students for whom English is an additional language, and often fail to detect well-disguised AI use.

Professional judgement and a clear understanding of students' typical work, rather than relying solely on detection tools, will always be the most effective way to identify potential misuse and avoid incorrectly accusing students.

Guidance from the Joint Council for Qualifications (JCQ) makes it clear that students must only submit work for qualification assessments that is their own. If they use AI in any way, it must be appropriately acknowledged. Failure to do so constitutes malpractice and may result in severe consequences, including disqualification.

Teachers and assessors must investigate suspected AI misuse and take appropriate action. The full JCQ guidance can be found in JCQ AI Use in Assessments.

Manual checks are one of the most effective ways to identify AI misuse. The JCQ outlines steps for educators in identifying misuse, keeping watch for inconsistencies in student work, and checking for unusual vocabulary, tone shifts and unverified references. If plagiarism or AI misuse is suspected, educators should try to confirm the source before reporting through the appropriate channels.

As we become better at discussing risks and spotting intentional and accidental misuse of AI, pupils and students are just as likely to become more and more adept at using generative AI, and so the likelihood of misuse may also increase.

There is the potential for this to happen during non-examination assessments (NEAs) at Key Stage 4 and 5 level. However, other less formal assessments take place throughout schools and colleges all the time.

These may be done under supervision or away from the classroom at home. Some pupils and students may be equipped with their own devices, whereas others rely on access to classroom-based PCs or laptops. Each context will be different, and how students and pupils access generative AI in support of their work in and out of school or college will require individual interpretation and should be reflected in your organisation's relevant policy.

While plagiarism is the uncredited use of someone else's work, AI misuse can involve over-reliance on AI-generated content in ways that bypass genuine learning. Schools and colleges must provide clarity on when AI is not to be used and promote use that doesn't bypass the development of skills and knowledge.

AI use is not just a concern at Key Stage 4 and 5, and students and pupils of all ages need to understand that their learning should come from their own engagement with the material. Embedding discussions about AI early helps students and pupils recognise that reliance on AI-generated content is not a substitute for genuine academic effort. It is important for students and pupils to understand that AI should not be used in place of independent thought and critical engagement.

We're still learning exactly how AI can best support student learning, but there's one principle that we need to keep in mind: if AI allows students to skip over the hard work of learning, it can actually harm their progress. This is reflected in findings from research in this area and also from cognitive neuroscience about how the brain builds lasting knowledge. Learning happens when students engage in what researchers call "desirable difficulties", these are tasks that require genuine mental effort and struggle. When a student wrestles with recalling information from memory or works through a complex problem step by step, or explains a concept in their own words, their brain is actively encoding and consolidating that knowledge into long-term memory.

Many researchers have likened it to building a muscle, the resistance is what makes you stronger. The same principle applies to learning. If AI becomes a shortcut that removes this productive struggle, it can actually prevent the formation of the neural pathways needed to learn something from forming in the first place.

Ben Whitaker and Steven Hope, Podcast Hosts, Edufuturists:

"Students are vulnerable to getting a quick first answer, because it gives you what you think you need – but then has the learning process of the struggle, of not knowing the answer straight away gone. That's part of what I think we need to make sure we wrap around – prompt engineering, AI literacy, these are all brilliant, but we need to start asking when should we use AI and when should we not? How do we know what we can trust?"

Presenter: For example, if a student asks AI to summarise a text they haven't actually read, or to solve a multi-step physics problem without working through it themselves, they're bypassing the exact cognitive processes that would help them learn. The AI might

give them the answer, but the student hasn't done the thinking that creates genuine understanding.

This is why we need to be thoughtful about when and how AI is used in learning contexts.

The Department for Education's updated generative AI product safety standards include measures to ensure the developers of AI products for education mitigate against the risk of "cognitive offloading", for example by providing responses that follow a pattern of progressive disclosure of information – starting with hints or partial steps, then gradually providing more detail. While these standards are intended to ensure that AI tools for education are designed to avoid undermining learning, educators should also take steps to ensure AI enhances learning – not replaces it.

Homework presents a particular challenge. AI tools can support educators in preparing learning materials, but it is not something that should be relied on by students and pupils to complete their work. Some may attempt to use AI tools to bypass the learning process, undermining their own development. Others may not yet understand that using AI-generated work could be a form of misconduct. Schools and colleges should set clear expectations and assignments should be designed to encourage independent thinking.

Daniel Emerson, Executive Director, Good Future Foundation:

"Academic integrity comes up a lot. Still, when we first started conducting research on what the state of play was with generative AI in schools back in 2023, 'will a student use AI to complete their homework' was the number one question that we were asked by teachers and also, you know, 'what can I do about that?' This still comes up very frequently. And, of course, if you're issuing homework that depends on a student completing a task, chances are they can find an AI that will complete that task for them. Most importantly, I think, for schools to consider is what does transparent use of AI look like. And are you fostering a culture of honesty and transparency around use? Because, let's face it, whatever these students do in the future, the likelihood is they're going need to be able to collaborate with this technology, and whatever job they go for, moving forward, we can see this happening already across so many different sectors. If schools take an approach of banning the technology, ignoring it, and not having informed conversations with students around academic integrity, and the benefits of using one tool, or one mode of a tool versus another, then the students aren't developing the skills that they're likely to need in the future, and you get more of this clandestine use of AI, where it's very, very difficult to detect."

Olivia Raven, Assistant Headteacher, Aylesbury High School:

"We've taken quite a balanced approach; we've looked not only at AI in CPD in terms of this is what it can do but we've also looked at what we call building an AI resilient curriculum. And we've thought about it more widely in the impact it might have on assessment and on homework. And given the staff the opportunity to discuss that, and to

plan with each other and discuss and share ideas with each other. And that will be ongoing because it's not going to go away."

Presenter: In the Chartered College of Teaching's Impact Journal special edition, an article by Lynsey Meakin on "AI and assessment: Rethinking assessment strategies" explores assessment and suggests how schools might rethink homework tasks they set for students in light of AI. Her key insight is that tasks requiring genuine critical thinking, problem-solving and analysis, where students must apply their knowledge in unique, personal ways, are far less likely to be completed entirely by AI.

The challenge schools and colleges now face is that traditional homework assignments can be easily completed using generative AI tools. When students can simply prompt AI to answer questions or complete written work, it becomes much harder for teachers to gauge what students have genuinely understood and learned themselves.

This is why Meakin emphasises that schools could favour homework tasks requiring personal reflection, individual reasoning, or practical application of knowledge. These types of activities are considerably harder for AI to complete convincingly because they demand the student's own thinking, experiences and unique perspective.

It is important to understand when discussing AI detection tools that they have limitations. False positives can unfairly penalise students, and AI detection tools are not reliable in distinguishing human and AI-generated work. Schools and colleges should develop guidance on how to approach such cases fairly, ensuring students are not disadvantaged. The focus should remain on ensuring students and pupils develop their own understanding and skills without relying on AI-generated work.

In written work, you can encourage citation of sources, and in other creative and technical subjects, you can focus on activities that make the process and iterations of the work visible and transparent. Most of all, it's important that students and pupils understand whether the use of AI constitutes cheating. Mock or practice assessments are a good opportunity to practise completing authentic work. If malpractice is suspected, discussing the work with a student can help as part of a holistic approach to investigating.

As new technologies and AI tools emerge, determining their acceptable use will be an ongoing challenge. Making clear the expectations around how much or how little students and pupils can use generative AI for their work, particularly in assessments, is important.

Any access to AI for students needs to be carefully planned, considering the DfE's policy position on the use of AI and its product safety requirements. Access must also align with Keeping children safe in education guidance.

Most freely available tools would likely not be suitable for student use, as they will not have the appropriate safeguards in place, and the model may learn from the prompts and information entered into them.

These conversations need to remain open and ongoing as technologies develop. Always ensure that you follow your setting's policies and guidance on this. By fostering a culture of integrity and setting clear expectations, schools and colleges can ensure AI does not compromise authenticity or bypass the important thinking that is part of the learning process. When students and pupils understand both the risks and ethical concerns of AI, they are better equipped to engage with learning in a way that prioritises their own effort and development.



Department  
for Education

© Crown copyright 2026

This publication is licensed under the terms of the Open Government Licence v3.0, except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://nationalarchives.gov.uk/doc/open-government-licence/version/3).

Where we have identified any third-party copyright information, you will need to obtain permission from the copyright holders concerned.

About this publication:

enquiries [www.gov.uk/contact-dfe](https://www.gov.uk/contact-dfe)

download [www.gov.uk/government/publications](https://www.gov.uk/government/publications)

Follow us on X: [@educationgovuk](https://twitter.com/educationgovuk)

Connect with us on Facebook: [facebook.com/educationgovuk](https://facebook.com/educationgovuk)