



ISSUE 1 | THE WALKER SHIELD - ONLINE SAFETY NEWSLETTER | 25.02.26

Welcome to our first edition of our new newsletter *The Walker Shield*, specifically designed to help you and your family navigate the ever changing online world and ensure that while we 'stay connected', we also 'stay protected'!

Each term we plan to cover a different online topic; explaining what it is, the benefits and threats, and how we as a community can keep ourselves and our young people safe as they surf.

This edition is a guide to...



ARTIFICIAL INTELLIGENCE



WHAT IS ARTIFICIAL INTELLIGENCE?

AI is a growing part of young people's digital lives, offering educational and practical benefits. However, it also brings concerns about misinformation, privacy, fairness, and safety.

AI is a branch of computer science focused on creating smart machines that perform tasks, usually requiring human intelligence. Like a person learns from experience, AI systems learn by analysing vast amounts of data to recognize patterns, solve problems, and make decisions.



Collects the dataset



Learns via neural networks and encodes patterns



Generates a new image



Refines the output

HOW DOES IT WORK?

AI isn't magic; it is **maths and code** working together.

→ **Training:** Engineers show the computer thousands of examples. For instance, to teach an AI to recognize a cat, they show it thousands of pictures of cats.

→ **Learning:** The computer finds common traits in those pictures (like pointy ears or whiskers) so it can identify a cat in a new picture it has never seen before.

→ **Improving:** The more examples the AI sees, the smarter and more accurate it becomes.

WHERE DO WE SEE AI?

You likely use AI every day without realising it. Here are a few common examples:

- **Video Games:** When you play against a computer opponent that reacts to your moves, that is AI.
- **Streaming Services:** When a movie app suggests a film you might like based on what you watched before, AI is making that recommendation.
- **Voice Assistants:** When you ask a smart speaker to play music or set a timer, AI is understanding your voice and responding.

Image and video content that has been created using AI, often pretends to be real and factual.

They can range from fake news clips to 'deepfake' abuse footage.

This content is often created to provide 'evidence' when the creators are trying to spread a false narrative. They can also be used to manipulate, intimidate, or even blackmail a victim.



BBC

iPLAYER



NETFLIX

Google Assistant

Siri



RISKS OF AI



EMOTIONAL MANIPULATION

GenAI content is often designed to strongly affect how the viewer feels. It is made digitally to guide the viewer's reaction, often to encourage donations, change opinions, or just to get views and likes. Examples include videos of people crying, hurt animals, or accidents. This kind of content can sometimes make children feel upset or that they can't do anything to help.



BLURRED REALITY

Seeing lots of fake things online can make it harder for children to trust what they see or hear. This is called the 'Liar's Dividend' - where something real is ignored because people think it's just 'fake news'. If children worry they won't be believed, they might not feel safe telling an adult about a problem.



DISINFORMATION

This is when untrue or made-up information is shared as if it were real, using AI to create material that looks very convincing and is designed to trick people. It is made so well that it can be hard to tell what is true and what is false. Examples of this include: made-up news stories, 'deepfake' interviews, and/or videos that seem to show a real event. If we don't learn how to understand media well (digital literacy), it's easy to believe that this fake content is giving us the facts.



IMPERSONATIONS & SCAMS

AI technology can mimic a person's look or voice using minimal photos or recordings. While this has creative uses, it also risks misuse, enabling fake videos of people like children, teachers, or celebrities. These impersonations can cause fraud, bullying, or embarrassment, spreading rapidly via social media or group chats. Young people especially must be aware of these risks and engage with digital content carefully.



PRIVACY & SECURITY

Some AI tools may request private information or track user activity. Incorrect privacy settings could expose personal details. Teach children good online habits and use strong privacy settings to ensure their safety.



BULLYING

Some people are using deepfake AI technology to create fake videos of their classmates that can be embarrassing or hurtful. These videos can be made to look like a student did or said things that they really didn't. Once these videos are shared, it can be very hard to get rid of them completely. This kind of online bullying, where technology is used to hurt others, is a serious issue and can be very challenging for those who are affected.



DEEPPAKE ABUSE CONTENT

Worryingly, some people are misusing AI to create harmful images of children. These pictures might be made from a real photo that was shared online innocently, or they could be completely made up by an AI program.

Making or even keeping these kinds of pictures is against the law and can cause long-lasting distress to the children involved.



DIGITAL LITERACY TIPS!

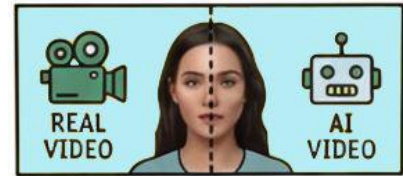
TEACH HOW IT WORKS



Explain that AI is in games, streaming services and even in programmes used in school. It uses patterns from old, existing information, to make choices and decisions. Because it isn't alive and doesn't think for itself or have feelings, some of the information it creates isn't always right. It may be biased due to the information it is learning from.

Use real-word examples like Netflix, Spotify, Youtube etc.

NOT ALWAYS WHAT THEY SEEM



Help children understand that some videos/images are created using technology.

Show comparisons of real and AI-made videos. Discuss how machines can imitate faces, voices, and actions. Learning about this technology builds confidence for safely navigating the digital world.

TALK ABOUT TRUST



Keep communication open.

If they find something unsettling, discuss it without judgement.

Create a safe environment for questions and be as honest as is appropriate.

This fosters trust and allows prompt action to be taken if necessary.

ENCOURAGE CRITICAL QUESTIONS



Be digital 'detectives', not passive viewers.

Ask: *Who made it?*

Why (to teach, laugh, or encourage you to buy)?

Is it trying to make you think a certain way?

Remember, even content sent by friends, may not be true!

Pause and wonder about unrealistic or shocking content.

STRENGTHEN PRIVACY SETTINGS

AI systems learn from vast amounts of data, sometimes including personal details. It's vital to be cautious about what you share online and understand the importance of privacy.

Model good habits: review app permissions and information collected by voice assistants (like Alexa, Siri). Keep voice notes, photos, and videos private.

Be aware that AI can create "deepfakes" using social media content. Teach young people to use privacy settings, limit public sharing, and turn off facial recognition to keep their information safer.



6 WAYS TO SPOT AI-GENERATED IMAGES



GLOSSARY

Algorithm

An algorithm is a step-by-step instruction set computers use for tasks like calculations or data analysis. For instance, an algorithm notices videos you like and suggests similar content, possibly from the same creator.

Algorithm Bias

AI systems can sometimes be unfair, just like a game that has been rigged! This unfairness can creep in because of the information used to teach the AI or because of choices made when building the AI rules.

Deepfakes

Pictures and videos that are deliberately altered to generate misinformation and disinformation. Advances in generative AI have lowered the barrier for the production of deepfakes.

Disinformation

Deliberate creation and spreading of false and/or manipulated information that is intended to deceive and mislead people, either for the purposes of causing harm, or for political, personal or financial gain.

Generative AI

AI that generates content including text, images, and audio in response to a user's prompts.

General-Purpose AI

AI models that can be adapted to complete a wide range of applications. This is the opposite of Narrow AI.

Hallucinations

Large language models, such as ChatGPT, are unable to identify if the phrases they generate make sense or are accurate. This can sometimes lead to inaccurate results, also known as 'hallucination' effects, where large language models generate plausible sounding but inaccurate text. Hallucinations can also result from biases in training datasets or the model's lack of access to up-to-date information.

Misinformation

The **unintentional** spread of false information. This can often be caused due to AI learning from incorrect or out of date data sets.

Narrow AI

Sometimes known as Weak AI, these types of AI are designed to perform singular, specific tasks (like speech recognition). They can not be adapted to complete other tasks.

Neural Networks

A neural network is a method in AI that teaches computers to process data in a way that is inspired by the human brain. It uses interconnected nodes or neurons in a layered structure that resembles the human brain. ChatGPT is a very large example of this.