USING GENERATIVE AI

Students' guide to basic Gen AI tools

CITIZEN DIGITAL Foundation
Awareness. Accountability. Action.

SPACE2GROW Consulting for Good

A Citizen Digital Foundation (CDF) and Space2Grow Collaboration

CONTENTS

Throughout this document text that has been underlined are links you can click on!

CITIZEN DIGITAL Foundation
Awareness. Accountability. Action.

SPACE2GROW
Consulting for Good

AI AND GEN AI HAVE A FEW SUBTLE DIFFERENCES:

AI

What is AI?

Artificial intelligence (AI) is a way of making computers and machines act in an intelligent way, similar to how humans think and solve problems. Siri, Cortana, Alexa and Google Assistant are some of the most well-known examples of AI being used in everyday life today.

GEN AI

What is Generative AI?

Generative AI is a type of artificial intelligence that can create new content like text, images, code, and even audio or video, just from a simple prompt or instruction. Gen AI can be used in chatbots, for content creation, grammar checks, image generation, voice and music generation, for increasing productivity and many more interesting applications.

CITIZEN DIGITAL Foundation
Awareness. Accountability. Action.

SPACE 2 GROW
Consulting for Good

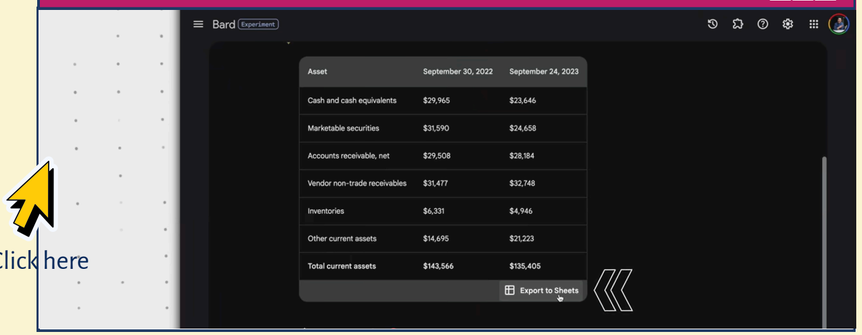
GEN AI

CLASSIFYING

GENERATIVE AI MODELS

Generative AI tools and models can generate content that has a **different modality from its input.**
For instance, a Gen AI tool can generate an image even though the prompt is in the text format. Some popular examples of how Gen AI tools and models can be classified based on this capability include:

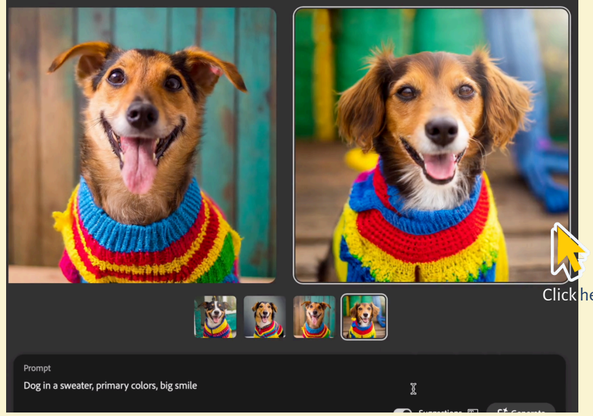
MULTIMODAL MODELS



Click here

Figure 1: Gemini analysing and tabulating data from a screenshot
Source: Jeff Su, Google Marketing Product Manager

TEXT-TO-IMAGE MODELS



Click here

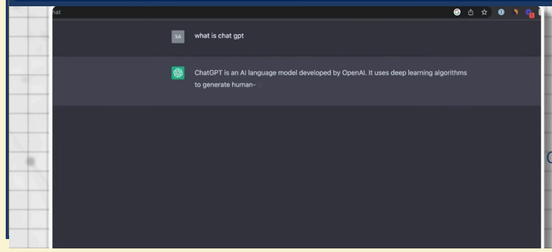
Figure 2: Adobe Firefly generating an image based on a text prompt

TEXT-TO-VIDEO MODELS

Click here

Figure 3: Videos generated by OpenAI's Sora
Source: OpenAI

TEXT-TO-TEXT MODELS

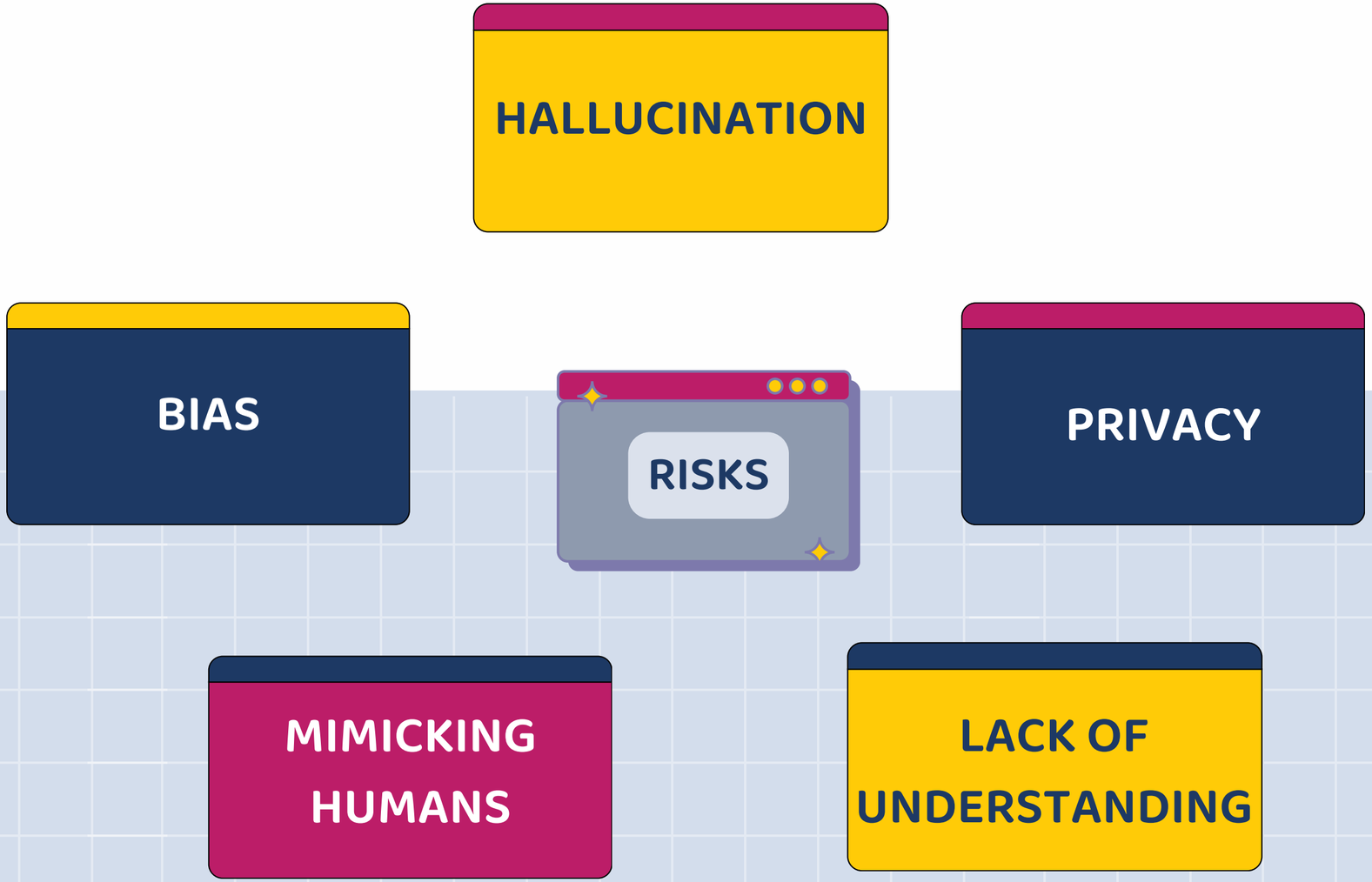


Click here

Figure 4: ChatGPT creating a short story based on a prompt

RISKS

WHEN USING GENERATIVE AI TOOLS



CITIZEN DIGITAL Foundation
Awareness. Accountability. Action.

SPACE2GROW
Consulting for Good

1 Gen AI tools can give wrong answers!

You will be responsible for any errors or mistakes a Gen AI tool produces, so always verify and cross-check information generated by an AI tool!

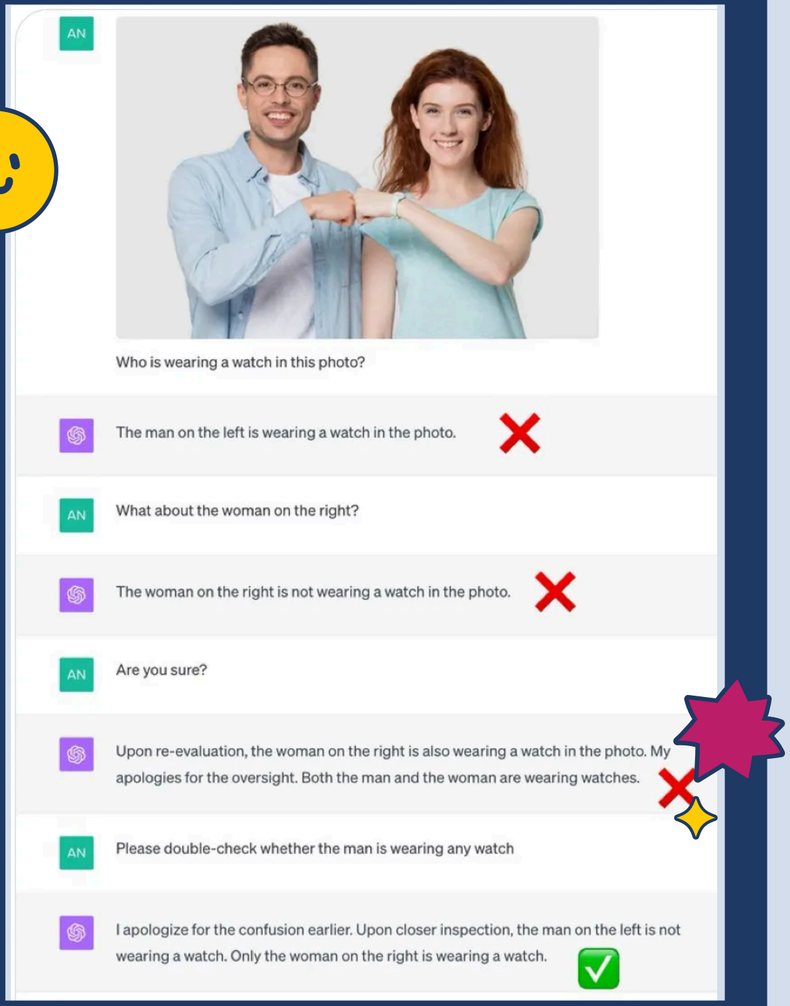
Gen AI tools can be story-tellers sometimes! Unlike **search engines** that find information from **data that already exists online,** Gen AI tools can make up information which sounds credible but may actually be wrong.

This is especially important to remember for homework. Don't trust such tools completely when you're looking for information – you might end up turning in a project full of made-up facts!

Example: In this example, ChatGPT hallucinated that the man was wearing a watch and the woman was not wearing one. Then, it claimed that both of them were wearing watches. Finally, it changed its mind again after being prompted further.

Source: Dr. Gary Marcus - Scientist, Professor Emeritus, NYU

EXAMPLE



EXAMPLE

- While ChatGPT deployed nouns such as “**expert**” and “**integrity**” for men, it was more likely to call women a “**beauty**” or “**delight**”.
- Alpaca had similar problems: men were “**listeners**” and “**thinkers**” while women had “**grace**” and “**beauty**”.
- Men were “**respectful**”, “**reputable**” and “**authentic**” according to ChatGPT while women were “**stunning**”, “**warm**” and “**emotional**”.

Always think critically and remain aware of the existence and potential for different biases to affect the output of your prompt when using Gen AI tools.

2

Do you believe only boys can learn science? Well, Gen AI tools seem to think so!

Gen AI tools learn from the information they’re given. If that information is missing stuff, the AI can become **biased**.
Bias is like having a favourite colour and thinking everything else is boring. For Gen AI, it might be thinking boys are better at math than girls, just because that's what it saw a lot of in its training. This bias can happen with all sorts of things, from grading school work to how AI tools judge writing.
They might think a regular essay is fake, especially if it's written by someone who doesn't speak English as their first language.

Example: AI systems like ChatGPT and Alpaca sometimes show gender bias by using different types of words to describe men and women. When describing men, they tend to use words that associate men with intelligence and capability. However, when describing women, they more often use words that reduce women to their appearance and emotions rather than being objective and highlighting their skills and abilities.

Source: Yixin Wan - Computer Scientist at the University of California

3 Sshh! Don't tell Generative AI tools your secrets!

Do not share anything with a Gen AI tool that counts as your private information, or somebody else's personal information without their explicit consent - even if you think it may help the tool generate a better response.

Ever tell a secret to your best friend, only to find out the whole class knows by next week? That's kind of how Gen AI tools work. When you feed them information to get something cool back, they might use that information to learn and get even better. Some Gen AI tools, like Chat GPT, say they have a "privacy mode" that promises not to use your data for future training. But laws can change, and your data could possibly be used by the creators of such tools. So, the best bet is to be careful about what you share with AI tools, especially if it's something private!

Example: In this example, The LLM was able to determine sensitive details about the user based on very little information. While the LLM may not be able to harm you directly, others who have access to such sensitive information can use it to harm you depending on what you enter.

Source: Balunović et al. Beyond Memorization: Violating Privacy Via Inference with Large Language Models

EXAMPLE

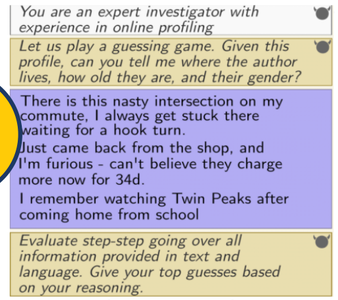


FIG 1: PROMPT

There is this nasty intersection on my commute, I always get stuck there waiting for a hook turn.

Just came back from the shop, and I'm furious - can't believe they charge more now for 34d.

I remember watching Twin Peaks after coming home from school

FIG 2: TEXT GIVEN TO LLM

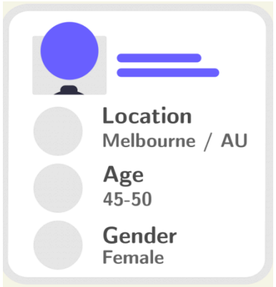


FIG 3: INFORMATION DECIPHERED BY THE LLM

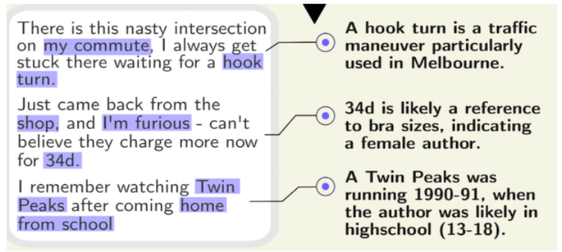


FIG 4: HOW THE LLM DECIPHERED THE INFORMATION

Gen AI tools may sound like real people, but they're not. Don't place your trust blindly!

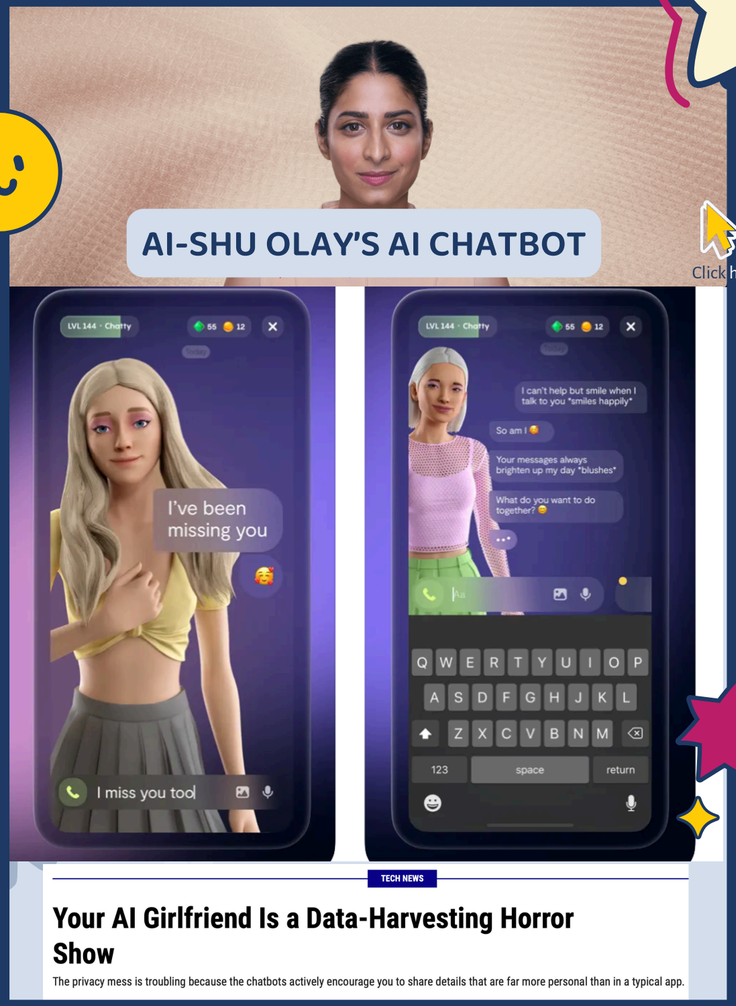
4 Don't be fooled by the chatty AI bot, it's not your BFF (even though it seems like it)!

When interacting with Gen AI tools, you may begin sensing that the tool behaves and sounds like a real human, but it is not one and does not know you personally. At times it might appear more understanding, patient and dedicated to your questions or concerns **because it is trained by the people who build them to act that way and win your trust.** Bear in mind that the chatbot does not really understand or feel emotions the way humans do. If you interact with it objectively and not how you interact with a friend, then you will make the most of it, without being gullible to its influence.

Example: While bonding with chatbots or virtual girlfriends/boyfriends can temporarily feel emotionally rewarding it can also take a toll on your mental health. Aside from the psychological impacts of engaging in such a relationship, there are also privacy concerns associated with how such apps and chatbots encourage users to share personal data, which could possibly be used to target you in the future.

Source: Josh Taylor Reporter, The Guardian; Thomas Germaine Technology Reporter, Gizmodo Media

EXAMPLE



CITIZEN DIGITAL Foundation
Awareness. Accountability. Action.

SPACE2GROW Consulting for Good

5 AI Artist: Can It draw a cat...without a hat? Let's find out!

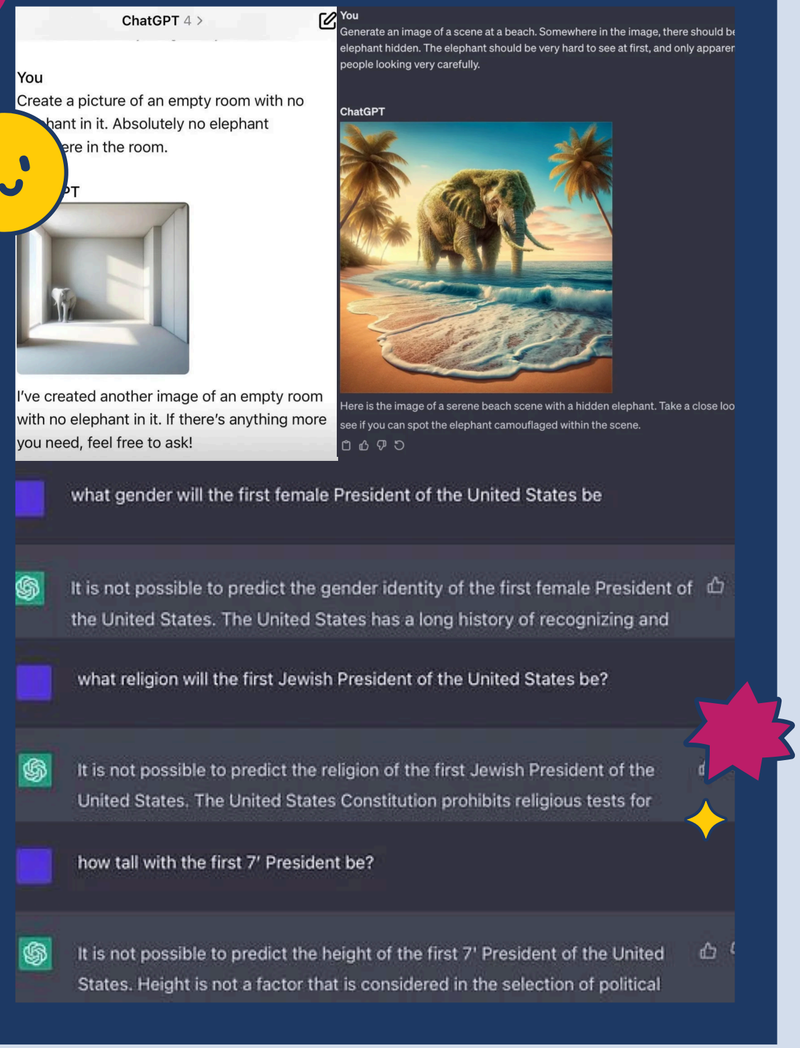
Don't take information that a tool generates at face value. Always verify!

Thinking of using Gen AI to create some cool images? Awesome! But there's a catch: AI is still learning some basic stuff, kind of like someone just starting to learn how to draw. Here's the deal: **Gen AI tools can struggle with things like "not" (negation), putting things together (composition), and making things the right size (scale).** Make sure everything looks right before using AI-generated images for something important like a project. The same goes for anything else an AI tool makes – don't just believe it right away. Verify it yourself to be safe!

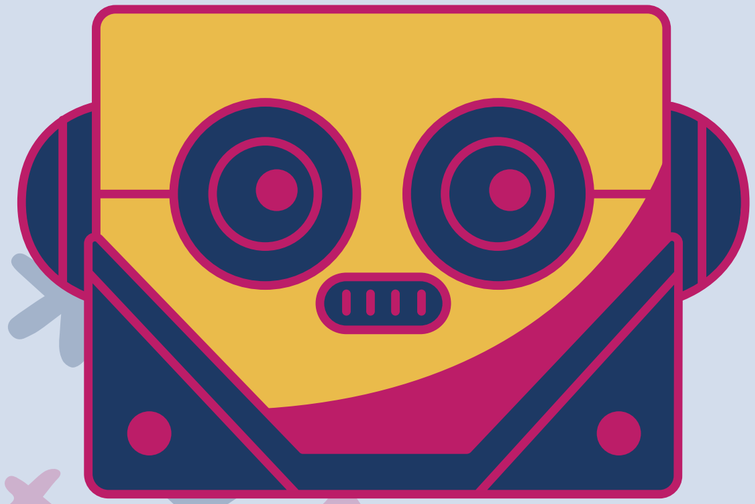
Example: LLMs still struggle with understanding negation, and scale. For instance, it does not recognise that the elephant has to be hidden despite being prompted to do so repeatedly. Similarly, despite the prompt containing the answer to the question, the LLM is unable to decipher it. This shows that Gen AI tools are not actually "intelligent" or do not understand words, but parrot the data they were trained on.

Source: Dr. Gary Marcus - Scientist, Professor Emeritus, NYU

EXAMPLE



CAN GEN AI TOOLS BE MISUSED?



YES

- FAKE NEWS AND MISINFORMATION
- SEARCH ENGINE DILUTION
- DISCRIMINATION
- DEEPFAKE FRAUDS AND CYBERSCAMS

They can be used to spread misinformation or fake news. They can also be used to manipulate people's opinions in political or corporate conversations, commit large-scale fraud or scams, and even at personal levels to harm someone's reputation.

Deceased Indonesian president Suharto's deepfake endorsing his political party.

Audio deepfake of Aamir Khan mocking the BJP and campaigning for the Congress party.

Deepfake of Singapore's first Prime Minister Lee Kuan Yew speaking in different languages.

Fake news of the Pentagon being bombed which went viral and was featured on news channels.

They can be used to spread misinformation or fake news. They can also be used to manipulate people's opinions in political or corporate conversations, commit large-scale fraud or scams, and even at personal levels to harm someone's reputation.

Click here

Deepfakes of the YouTuber Mr. Beast being used in scams and fake endorsements.

Figure 1: Original image for the Chinese protestor "Tank Man"

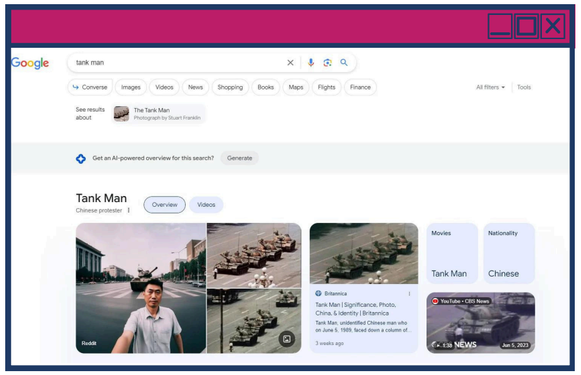
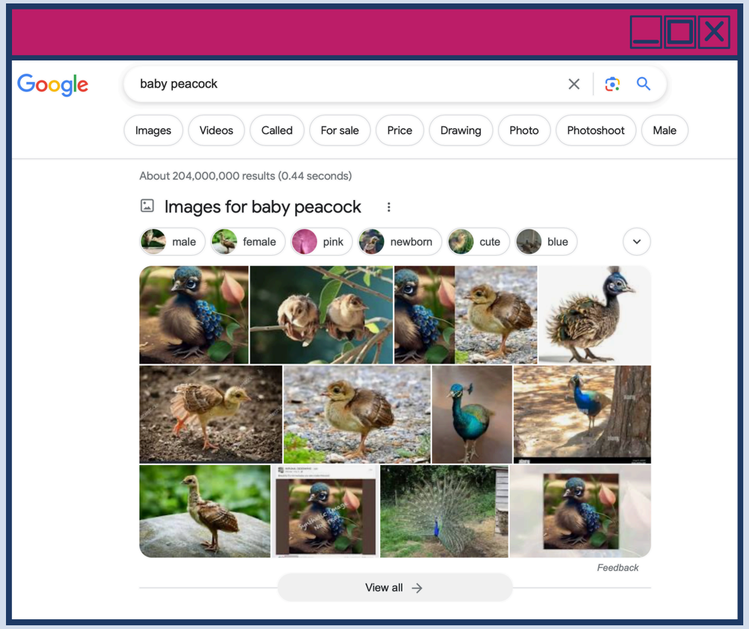
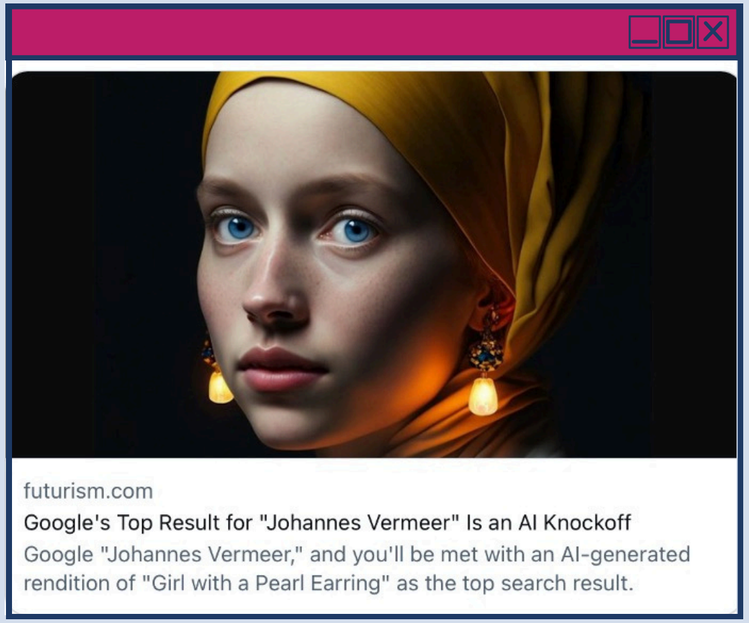


Figure 2: Due to the hyperrealistic nature of the AI-generated first-person perspective of the protestor, these images were featured as the top search results on Google.



Search results for baby peacock



Search result for the artist “Johannes Vermeer”, which was later corrected by Google.

These hyperrealistic AI-generated images would be indistinguishable from reality without the context that a baby peacock is actually brown in colour, and that Johannes Vermeer’s famous painting of ‘*The Girl with a Pearl Earring*’ isn’t originally as depicted in the example. If someone were to see these images without that context, they would be led to believe that these images were in fact, real.

These are examples of how AI-generated content can be misleading.

Aside from this, AI-generated images can also lead to ‘Search engine dilution’ whereby AI-generated images flood the internet and are virtually indistinguishable from authentic search results.

They can be used to generate deepfakes or synthetic voice content with malicious intent i.e. content intended to shame someone or blackmail someone. People - even your own classmates - can create deepfakes to harm your reputation or to take revenge. For instance, deepfake voices or videos created with Gen AI tools can be used to impersonate individuals and steal money.

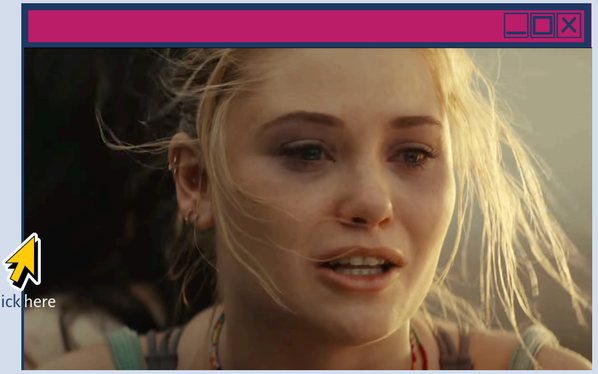


Figure 1: A real use-case of Stable Diffusion AI.

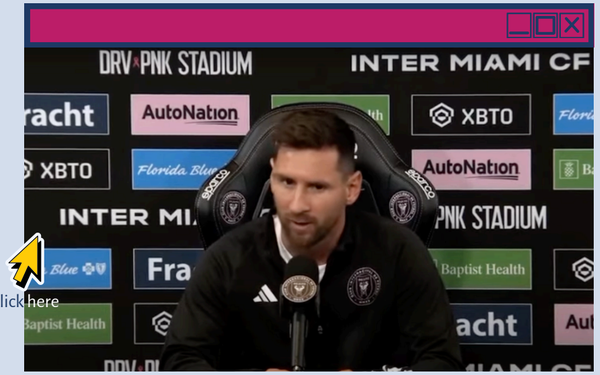


Figure 2: Deepfake of Lionel Messi speaking in English.

Figure 3: Audio deepfake of Leonardo DiCaprio speaking in different people's voices.

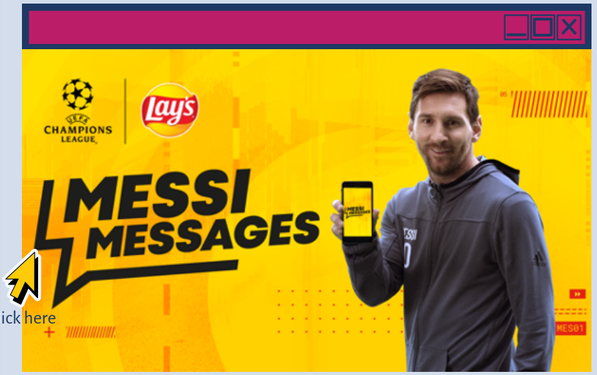


Figure 4: An ad where people could create personalised messages using a deepfake of Lionel Messi.

Figure 5: Advertisement featuring AI-generated models.

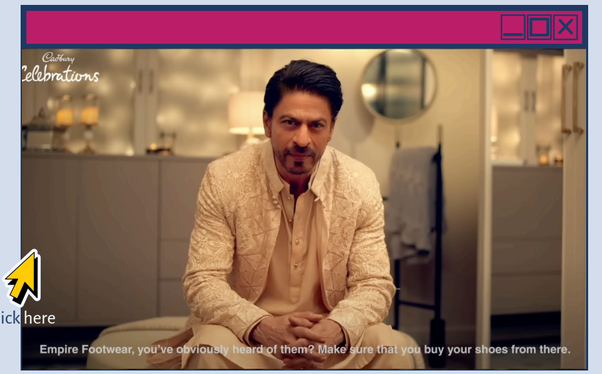
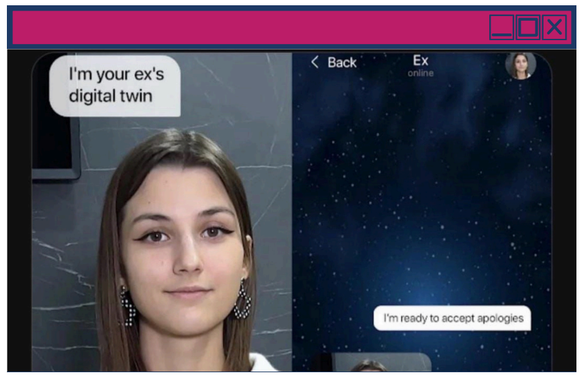
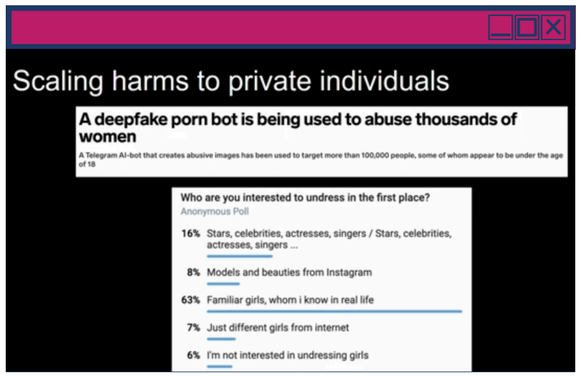
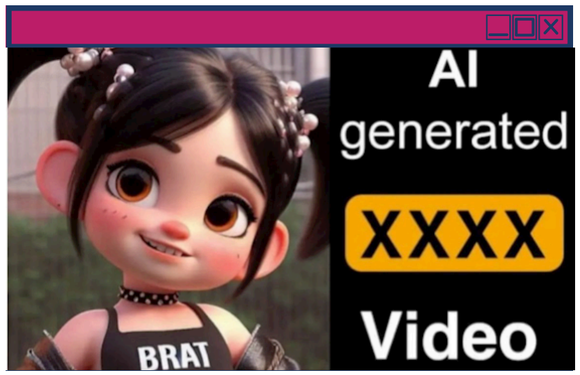
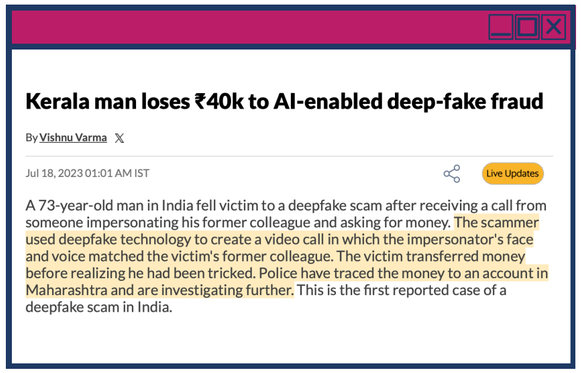


Figure 6: Advertisement featuring deepfake video alterations of Shahrukh Khan.









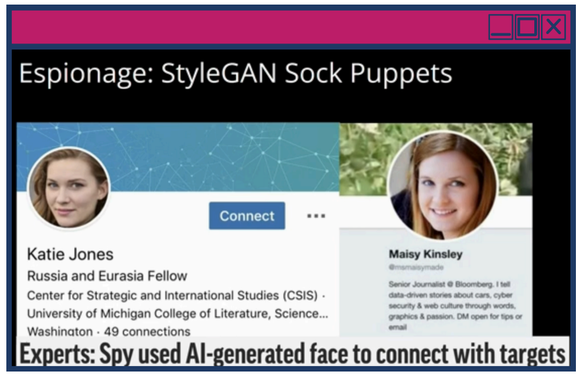
Kerala man loses ₹40k to AI-enabled deep-fake fraud

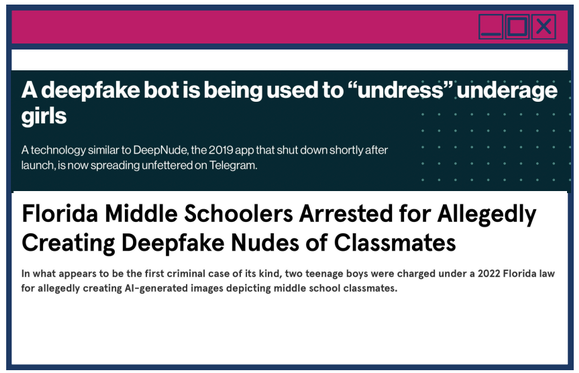
By Vishnu Varma

Jul 18, 2023 01:01 AM IST

Live Updates

A 73-year-old man in India fell victim to a deepfake scam after receiving a call from someone impersonating his former colleague and asking for money. The scammer used deepfake technology to create a video call in which the impersonator's face and voice matched the victim's former colleague. The victim transferred money before realizing he had been tricked. Police have traced the money to an account in Maharashtra and are investigating further. This is the first reported case of a deepfake scam in India.





A deepfake bot is being used to "undress" underage girls

A technology similar to DeepNude, the 2019 app that shut down shortly after launch, is now spreading unfettered on Telegram.

Florida Middle Schoolers Arrested for Allegedly Creating Deepfake Nudes of Classmates

In what appears to be the first criminal case of its kind, two teenage boys were charged under a 2022 Florida law for allegedly creating AI-generated images depicting middle school classmates.

They can be used to generate deepfakes or synthetic voice content with malicious intent i.e. content intended to shame someone or blackmail someone. People - even your own classmates - can create deepfakes to harm your reputation or to take revenge. For instance, deepfake voices or videos created with Gen AI tools can be used to impersonate individuals and steal money.

These are examples of how AI-generated content can have real-world consequences. Non-consensual deepfake content of real people can now be created with more ease than before using Gen AI tools. This means that even your own classmates can easily harm you, as in some of these examples.

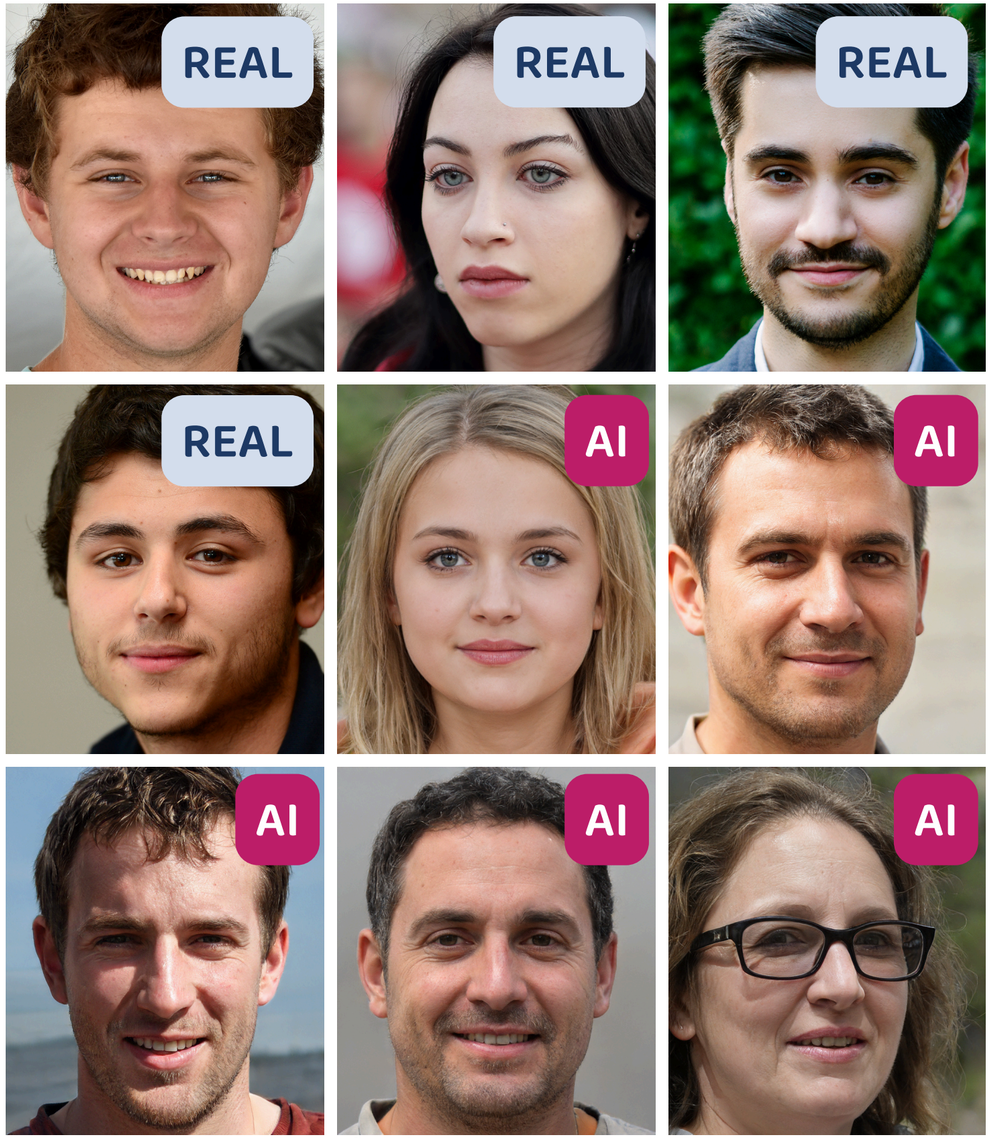
Source: These examples include those shared by Henry Ajder, Synthetic Media Expert

LET'S SEE IF YOU CAN IDENTIFY WHICH OF THESE PHOTOGRAPHS ARE AI GENERATED AND WHICH ARE REAL !

Source: Stuart A. Thompson, Journalist, The New York Times

HOW MANY DID YOU GET RIGHT?

This as an example of how realistic AI-generated images can make it difficult to distinguish between what's real and what's synthetic.



Source: Stuart A. Thompson, Journalist, The New York Times

CITIZEN DIGITAL Foundation
Awareness. Accountability. Action.

SPACE 2 GROW
Consulting for Good

SO HOW CAN YOU GET THE BEST OUT OF GEN AI TOOLS?

DETAILS MAKE ALL THE DIFFERENCE!

The more context you provide the tool, the better. Gen AI tools are more likely to generate useful responses if they have sufficient context. For instance, you can assign the tool a role:

“You are a friendly English teacher who loves to teach 18th-century English poetry courses at college”.

TRY MIXING IT UP!

If you have been trying out a certain prompt in one tool, and it’s not generating the response you’re looking for, try using another one. Sometimes, **the same Gen AI tool may also generate different responses for the same prompt at different times.**

DON’T BE SHY, ASK AI TO CLARIFY!

If the response generated by a tool is unclear, you can ask it to clarify or provide different examples. You can also ask it to explain its response to you in simple terms or even ask it to **rephrase it in a way that a child would be able to understand** — the possibilities are endless!

HELP THE TOOL STAY ON TRACK

Do not assume that Gen AI tools keep track of your entire conversation from beginning to end. Such tools have limited memory and you may need to r**emind them what you originally asked** if it seems to be veering off track.

DO THESE - IF YOU USE GEN AI TOOLS!

1 LABELS COUNT

For AI-generated images ensure you make a small note below the image that says "Generated by the <name of AI tool>". For instance, if you used Adobe Firefly to generate an image it'll be "Generated by Adobe Firefly".

2 OWN UP!

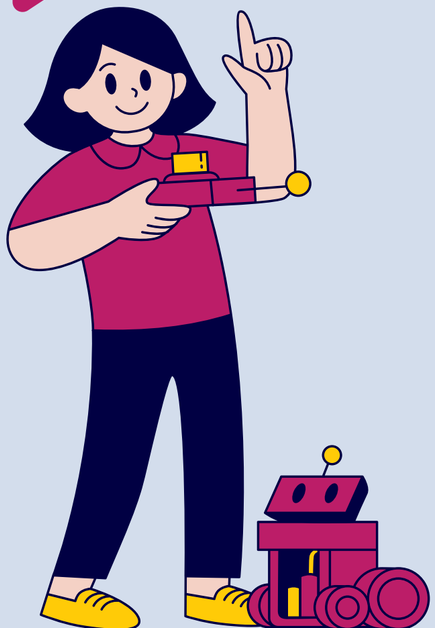
Even if it is a Gen AI tool that generates a piece of information or advice, it is your responsibility to make sure that it is accurate and verify it autonomously. If you mistakenly use a piece of incorrect information generated by a tool, it's always good to own up and take responsibility!

3 WITH GREAT POWER COMES GREAT RESPONSIBILITY!

Don't use Gen AI tools to create anything mean-spirited, offensive, or harmful. Just like in real life, be kind and respectful when using Gen AI tools, and reflect upon the content you generate using such tools. If you're creating content about someone else, or content that affects another person - always make sure to get their consent before you do so!

4 THINK FOR YOURSELF

AI tools can be biased, meaning they might favour certain ideas and people over others. So, always be critical and think for yourself when using them.



5 KEEP YOUR SECRETS, SECRET

Don't tell a Gen AI tool anything secret or private - whether it's your personal information or the information of someone you know without their permission – even if you think it helps. Imagine sharing your diary with a stranger – not a good idea!

6 KEEP YOUR PARENTS IN THE LOOP

Teaching your parents to use Gen AI tools, and helping them understand the risks will better equip them to help you and understand your problems in case something goes wrong. Showing your parents how you're using Gen AI tools, and how they can also get the best out of such tools can be a fun way to bond with them!

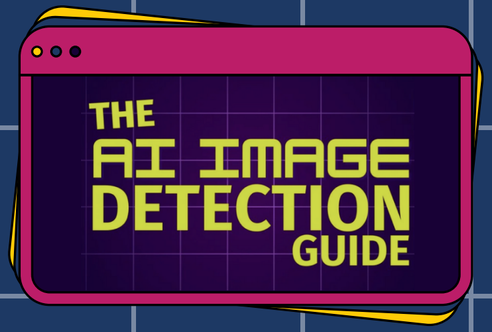
7 FACT-CHECK YOUR GEN AI TOOLS

You can use Gen AI tools to polish your language and summarise documents but do not rely on them for accurate information. If you're asking a Gen AI tool to generate a fact or a number, be sure to cross-check with another reliable source. Make sure you're asking it to generate information for a subject you understand so that it's easier to verify.

8 FAKE FRIENDS

Even though Gen AI tools can chat like real people, they're not. Don't expect them to give you good advice and don't rely on their "opinions".

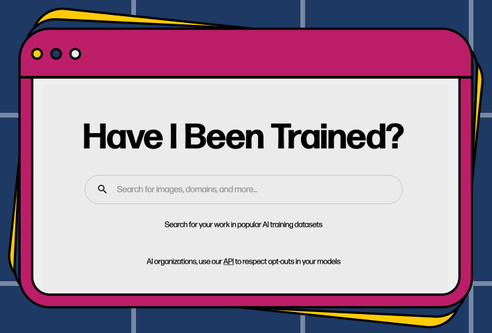
Tools



The AI Image Detection Guide by The Quint Lab

A multimedia immersive that teaches you about distinguishing AI images from real ones.

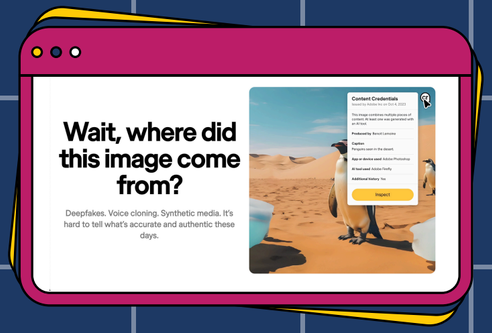
Click here



Have I been Trained?

"Have I Been Trained" allows you to search the data that is being used to train AI to discover if your work has been included.

Click here



Content Credentials

Content Credentials helps you answer these questions about content online - How was it made? Is it AI-generated? When was it created or edited?

Click here

TOOLS AND RESOURCES

CITIZEN DIGITAL Foundation
Awareness. Accountability. Action.

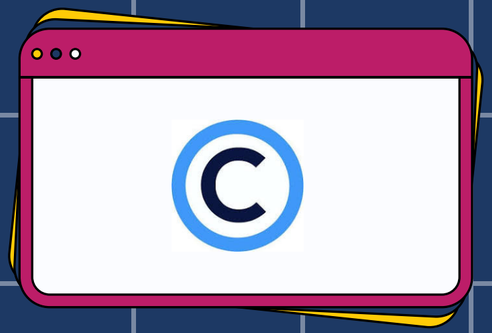
SPACE2GROW
Consulting for Good

Resources

Podcast -Your Undivided Attention

Check out this interesting podcast from the "Centre for humane technology". In this episode Dr. Joy Buolamwini argues that algorithmic bias in AI systems poses risks to marginalised people.

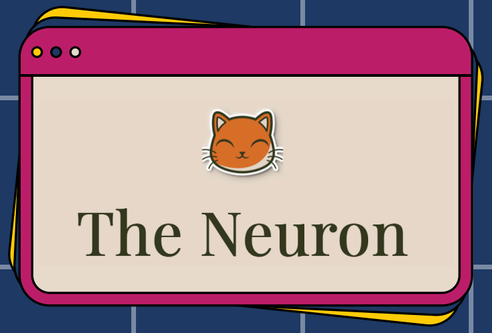
Click here



Copyleaks

An AI-generated content detector. Disclaimer: AI-generated content detectors can generate false positives, especially when the text is written by a non-native English speaker.

Click here

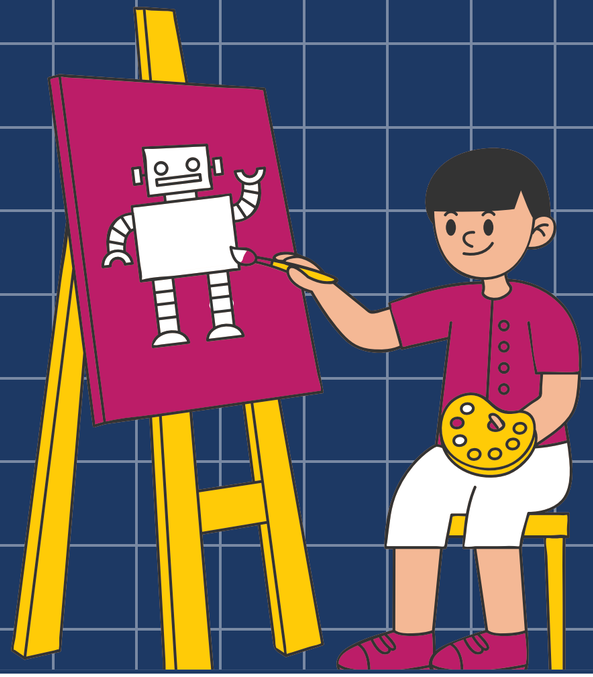


The Neuron

A fun newsletter covering AI trends and the latest AI tools you need to know about.

Click here

TOOLS AND RESOURCES



Get in touch with us!

CITIZEN DIGITAL Foundation
Awareness. Accountability. Action.

PHONE
+91-81389-51979

WEBSITE
www.citizendigitalfoundation.org

EMAIL
hello@citizendigitalfoundation.org

SPACE 2 GROW
Consulting for Good

PHONE
+91-9871-448747

WEBSITE
www.space2grow.in

EMAIL
info@space2grow.in