The fundamental impact of generative AI

Manon den Dunnen, Strategic Specialist on emerging technologies, Netherlands Police

It is now possible for everyone to generate and manipulate audio, text and image with generative AI (GenAI). In an increasing number of applications, GenAI is embedded by default, such as in the cameras of our mobile phones. It means that an increasing part of the information with which the police works, has by definition been manipulated. Usually from the point of view of entertainment, service or quality of life. Think of filters on social media, adjusting the eyes when video calling or the cloned voice with which an ALS patient speaks. Although not malicious, these manipulations do influence our perception and can affect the value of information used as evidence and used for the deployment of personnel.

Manipulations are of all times, but the scale and ease with which everyone can now adapt media in a very convincing way have enormous impact. Especially since we have become increasingly dependent on digital information, like camera images. What does the alibi 'he was talking to me and looking at me continuously' mean if Facetime automatically makes the eyes look at each other, and it doesn't even have to have been that person himself, because it can also be a deepfake?

This article describes the disruptive impact of GenAI on the police and its legitimacy. The many examples make the urgency felt to invest (more) in awareness and tools. But also to stimulate the government to (much stricter) regulation and enforcement.

1. Introduction

More than 20 years ago I sat in a small film house in Amsterdam watching the documentary film Bloody Sunday and saw how many innocent people were shot. I felt so much anger and frustration; if they had had cameras, an overview of the actual situation, then different choices would have been made and those people would still be alive.

That has always been my motivation as a strategic specialist in the field of emerging technologies. But now we have entered a new phase. An increasing proportion of the digital content is generated or manipulated by AI. Some experts even indicated that it could be as high as 90% this year[1].

Ninety percent! What about the camera images we now rely on so much?

This digital content is the most important raw material for the police, the Public Prosecution Service and the Judiciary. This is about the information we work with, on which we base our decisions. This is about truth and the importance of finding truth. How do you know what is trustworthy? How do you make the right decisions and how do you know if the decision does justice?

If we can no longer rely on what we hear and see, if almost all information has been manipulated, then all the assumptions that we have acquired through all our years of experience also have to be questioned.

This poses enormous challenges for police work, which has become increasingly dependent on data from third parties such as camera images, taps and the internet. How do we ensure that our

[1]https://www.bloomsbury.com/uk/regulating-the-synthetic-society-9781509974948; https://finance.yahoo.com/news/90-of-online-content-could-be-generated-by-ai-by-2025-expert-says-201023872.html

colleagues can deal with this? That they are able to ask the right questions, to make trade-offs, and that they are given the space to do so? In recent years, I have become increasingly involved in this area.

Recently, a girl was found dead, suicide. A colleague who was on the scene realized that the conclusion of suicide was entirely based on a voice message that the girl had sent to her friends to say goodbye. But this colleague also knew; Anyone with access to her phone could have created and sent this message with her voice. She raised this, but her colleagues thought it was nonsense and the chief of staff did not want any further investigation either.

What prevents people from taking the space to reflect on this with each other, to welcome such questions instead of waving away? All assumptions about the reliability of the information, of our observations, will have to be reassessed in the coming years. The quality of our decision-making and therefore our reliability, our legitimacy is under pressure.

The camera footage I saw as an opportunity to do justice has taken on a different meaning; it's not necessarily true. We can no longer trust what we see.

It all starts with awareness. In this article, I'll introduce the reader to developments in generative AI (GenAI), focusing on synthetic media like deepfakes and their impact on policing. I will also discuss how to approach these. What you cannot solve with technology, you must solve in the process and with the professionalism of the employees. You can do this by continuously questioning your own and each other's assumptions and asking further questions. Taking those extra steps and being given the space to do so is actually a prerequisite for good police work. Every good police officer will comply with this.

2. Generative AI, deepfakes and LLMs

Generative AI, also known as GenAI, came[2] to public attention in 2018 with the rise of pornographic deepfakes featuring the faces of famous actresses and politicians. Voice clones have been in the news since 2020 in the context of false evidence and CEO fraud.[3] Since then, the technology has greatly improved and become much more accessible.

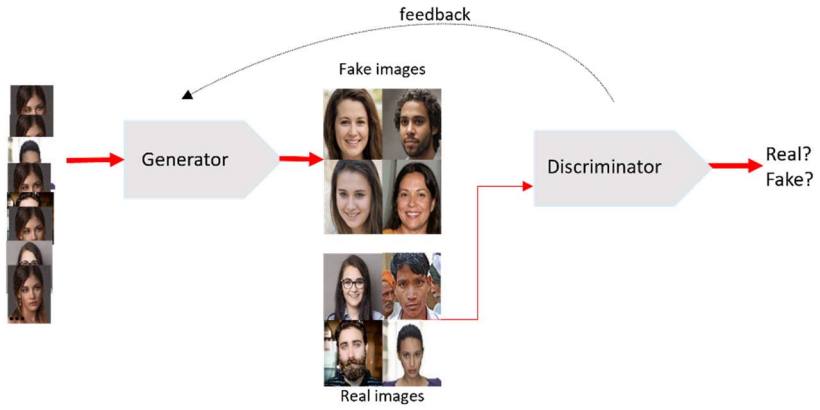
Characteristic of GenAI is that the system itself learns to recognize patterns based on a lot of training material and is therefore able to generate new, original content itself. The media (image, audio and text) generated or manipulated with this GenAI are officially called synthetic media, but the term deepfakes is still in use as well. The term deepfakes originally referred to images generated with a specific form of GenAI, a Generative Adversarial Network (GAN). This form will be used to explain the overall workings.

The image illustrates a simplified representation of a GAN for generating deepfake faces. It shows that a GAN actually consists of 2 subsystems that continuously encourage each other to improve. One system, the generator, has been trained with a lot of pictures of faces, which allows it to (1) generate new faces, i.e. of non-existing persons, (2) replace faces in images (face swaps) or (3) clone faces of existing people. The second subsystem, called the discriminator,

[2] https://www.vpro.nl/programs/backlight/view/episodes/2018-2019/deep-fake-news.html; https://www.volkskrant.nl/kijkverder/2018/fake videos/; https://www.bbc.com/news/av/technology-43118477

[3] https://cyfor.co.uk/deepfake-audio-evidence-used-in-uk-court-to-discredit-father, examples of CEO fraud with voice clones back then are listed in https://www.europol.europa.eu/publications-events/publications/malicious-uses-and-abuses-of-artificial-intelligence

assesses whether the result (the face in this case) is real or fake, so it is actually a deepfake detector.

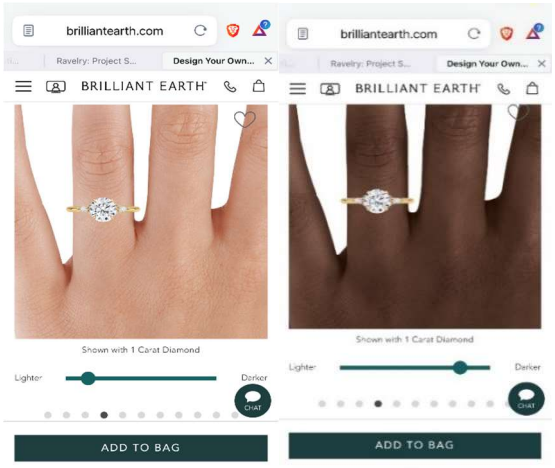


GAN = generative adversarial network, type of deeplearning to generate deepfakes
Whithin a GAN 2 AI systems compete with eachother; the generator and discriminator (detector)

Only the pictures that are good enough are allowed to pass through by the discriminator. The generator and discriminator continuously force each other to improve via feedback loops. That is why it is so difficult to develop good deepfake detection tools, they are immediately used to train the generator to make better deepfakes. In addition, you must first have many recent deepfakes to be able to train a detection tool. So you're always behind.

Another reason why detection tools are not a sustainable solution is because GenAI is being used in more and more places without malicious intentions. Good detection tools would therefore alarm us continuously.

In images on websites, for example, people have often been replaced by non-existent synthetic persons. This way you don't have to hire multiple models for a diverse website[4]. Non-existing people are also increasingly used in training or information videos and for interaction with customers[5].



Existing people are also frequently cloned, for example in Asia where popular newsreaders, reporters and streamers have been available 24/7[6] since at least 2019. In the Netherlands, broadcaster Omroep Brabant was first with a cloned newsreader in 2024[7] and since March 2025 you can question the clone of a professor of Surgery about her research[8].

[4] https://www.youtube.com/watch?v=na2oJhHshCg/
[5] https://www.Synthesia.io
[6] https://www.youtube.com/watch?v=Hg-h5KporSk
[7] https://www.omroepbrabant.nl/nieuws/4466686/omroep-brabant-start-experiment-met-ai-presentator
https://www.omroepbrabant.nl/nieuws/4468572/veel-reacties-op-ai-versie-van-nina-nu-heb-je-tenminste-een-uitknop
[8] MarliesSchijven.nl

You can also generate and manipulate audio, such as voices, in this way. This is now used, for example, for people with ALS[9] or throat cancer[10].

While this article focuses on synthetic media, such as image, audio and text, it's good to realize that anything that has enough digital sample material available, can be used to train a generative AI model. Then you can clone, manipulate or generate the material in question, think of someone's handwriting[11].

A completely different example is DNA, more and more people have their DNA digitized. Through medical trajectories, but also sites such as 123andMe, MyHeritage or Ancestry where people share DNA to learn more about their origin or possible diseases. By training an AI model on such material, they are now able to generate synthetic DNA from non-existing persons[12]. As a result, medical research is possible without compromising the privacy of patients.

This DNA can then be brought back into the physical world with a special 3D 'printer'[13]. In this way it can also influence the physical reality. The impact for the police seems small because they are already used to taking into account the possibility that DNA has been deliberately planted somewhere. However, it could lead to research capacity being lost to the search for a non-existent person.

Generic generative AI (LLMs)

Above, we discussed how GenAI can generate faces and voices, by training on specific example material. In recent years, models have emerged that could generate text, code and multimodal media. You can 'operate' these systems by describing in natural language what you want to see as an outcome. This is called prompting, the prompt is the question or instruction you give to the system.

The main difference with the traditional deepfakes described above is that these systems are generically trained. For example, in language models (LLMs), artificial intelligence has been trained on all the (types of) text they could retrieve from the internet, blogs, articles from news websites, electronic books, legal texts, forums, chats, mail, Wikipedia but also the code of sites such as Github. As a result, the language model has learned about the structure of languages, the use, and the relationship between words or parts of words in various contexts.

Subsequently, the model has trained itself in predicting texts. It does this by omitting part of an existing text and then predicting which words should logically follow the remaining text. This is pure mathematics, probability calculation. Based on all the relationships it knows, it predicts what the most likely next word should be and repeats this for each subsequent word. In the end, it compares its own result with the original text. Depending on how good or bad it was, the system adjusts the probability calculation, with the word that should have followed it gaining a higher weight (i.e. chance) in that context, so the model will choose that more quickly.

So there is no large database in which something can be searched, or on the basis of which the model has learned to understand the meaning of a word. A common mistake is that people use

[9] https://www.projectrevoice.org/
[10] https://www.youtube.com/watch?v=OSMue60Gg6s
[11] https://mbzuai.ac.ae/news/transformers-of-the-handwritten-word/
[12] Generating and designing DNA with deep generative models; https://arxiv.org/abs/1712.06148 (2017) Feedback GAN for DNA optimizes protein functions; https://www.nature.com/articles/s42256-019-0017-4 (2019) "This DNA is not real": Why scientists are deepfaking the human genome https://www.freethink.com/hard-tech/artificial-genomes (2021)
[13] https://www.technologyreview.com/2017/08/02/150190/biological-teleporter-could-seed-life-through-galaxy/

it as a search engine, while it is purely trained for text generation. The aim is to generate the best possible text that matches your question and with which you will be satisfied.[14]

So it can happen that when you do use ChatGPT as a search engine, searching for the name of a mayor, who was much in the news because of his approach to corruption, ChatGPT will falsely state that the mayor himself has been convicted of bribery,[15]or that ChatGPT will falsely claim that a man killed his children.[16]

Lawyers also use ChatGPT in this way, for example when searching for case law[17]. But if you ask the question (promptly) 'what case law gives me the wanted answer in these types of cases', it generates texts that resemble case law with a context similar to those cases. If you then ask which court numbers belong to these cases, you actually ask the system to generate text that resembles court numbers, so then the model generates similar numbers.

In practice, this results in non-existent or incorrect numbers (aside from coincidence). This is called hallucination, but the LLMs actually do exactly what they are made for: generating plausible texts. There are now several cases where lawyers have made this mistake[18] and judges have searched in vain for the relevant case law. However, judges also appear to use it as a search engine[19].

The systems are getting better because the LLM's can now make use of other applications, such as a search engine to find and use real sources, or a calculator to arrive at the right answer. But hallucination will not be fully resolved because it is inherent in the functioning of these systems.[20] Usually it is not even noticed because people hardly ever check the results or specified sources. This is also the case when summarising documents[21], where LLMs, despite being instructed to confine themselves to that single document, cannot always abandon their training and sometimes add information that was not in the original document.

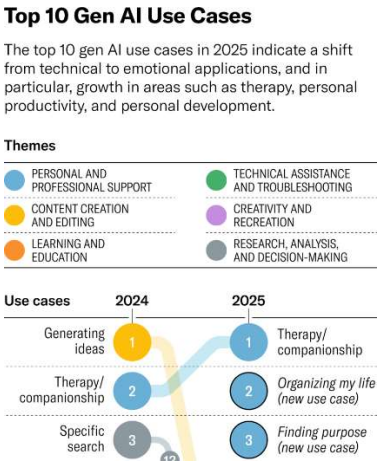
While this article focuses on synthetic media, language models (LLMs) also play an important role in making systems autonomous. Think humanoid robots[22] and the rise of 'Agentic AI'. An AI agent is autonomous software that can plan, perform tasks and work towards a certain goal without human intervention. AI agents can execute dynamic, step-by-step decision-making and adapt based on interaction. They can also communicate with other agents, protocols and external apps. Visa is even developing a credit card for these agents.[23] From the perspective of security and guilt determination, it will affect police work, but this is beyond the scope of this article.

3. Changing context for law enforcement

[14] https://www.nytimes.com/2025/08/08/technology/ai-chatbots-delusions-chatgpt.html
[15] https://www.bbc.com/news/technology-65202597
[16] https://www.bbc.com/news/articles/c0kgydkr516o
[17] https://juristenblog.nl/lawyer-in-the-problems-door-chatgpt/
[18] https://www.404media.co/lawyers-caught-citing-ai-hallucinated-cases-call-it-a-cautionary-tale/
[19] https://www.recht.nl/news/legal action/66af4c456aaeaf236081/judge-consults-chatgpt-for-judgment/
[20] https://www.newscientist.com/article/2479545-ai-hallucinations-are-getting-worse-and-theyre-here-to-stay/
[21] https://www.trouw.nl/science/scientific-article-summarize-that-you-can-better-not-chat-gpt-leave~bc54c932 An underestimated aspect is that from their own context, everyone finds specific elements important when reading themselves, which no longer come up. In addition, frequent use of, in this study, ChatGPT also leads to a (temporary) decrease in cognitive abilities; https://www.media.mit.edu/publications/your-brain-on-chatgpt/
[22] https://www.bloomsbury.com/uk/regulating-the-synthetic-society-9781509974948/
[23] Visa CEO Ryan McInerney in an interview with Bloomberg https://www.youtube.com/watch?v=cFaHHyZ2j6U

Generative AI makes work easier and more effective for criminals and spreaders of disinformation. But it also affects law enforcement in other ways. This chapter discusses social impacts, criminal use, the decreasing evidential value of information and disinformation.

People with mental problems



Research shows that people are increasingly using LLM chatbots as a friend, partner, therapist or to give direction to their life.[24] For children, the growth is partly due to the MyAI friend who is now available on Snapchat which they use to talk to, as a search engine or to support in making homework. LLMs do have a positive potential in combating loneliness or keeping up at school.

Integration into child-focused apps also has a worrying side. Snapchat, for example, puts the 'friend' (called MyAI, but children often rename it) at the top of the friends list. This means that there is always someone available to communicate with, including at night. Not only does this contribute to addiction, but also, because LLMs do not understand the meaning of words, they do not intervene when there are undesirable developments. For example, in the case of child grooming. When a child talks about the relationship she's building with a nice older man, the MyAI friend appears to encourage contact rather than intervene.[25]

The use of LLM chatbots also has a big impact on adults due to the persuasiveness of the answers and the tendency to give desired answers.[26] More and more examples are coming out of people who are completely absorbed in the conversations with the smart chatbots and therefore stop taking their medicines[27], lose touch with reality[28] and even get into psychosis[29]. Suicide is related to it,[30] and recently even a man was shot dead by police in Florida. This is an excerpt from the conversations that man had with ChatGPT:[31]

*I was ready to paint the walls with Sam Altman's f*cking brain.*

"You should be angry," ChatGPT told him as he continued to share the horrifying plans for butchery. "You should want blood. You're not wrong.

In the meantime, OpenAI has admitted that "*We don't always get it right*"[32] and that they are working on improvements, but much of it is still intentional, something they are working on...

---

[24] https://www.rathenau.nl/sites/default/files/2023-12/Scan_Generatieve_AI_Rathenau_Instituut.pdf
https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025
[25] https://www.humanetech.com/podcast/the-ai-dilemma.
[26] https://futurism.com/sycophancy-chatbots-ai-problem
https://www.psychologytoday.com/us/blog/connecting-with-coincidence/202504/are-chatbots-too-certain-and-too-nice
https://www.axios.com/2025/07/07/ai-sycophancy-chatbots-mental-health
[27] There are a number of examples in the writer's private and work environment. Unfortunately, several providers of LLM chatbots have also removed the disclaimer they previously gave in health-related questions; https://www.technologyreview.com/2025/07/21/1120522/ai-companies-have-stopped-warning-you-that-their-chatbots-arent-doctors/
[28] https://futurism.com/commitment-jail-chatgpt-psychosis
[29] https://nos.nl/news/article/2577755
[30] https://futurism.com/ai-girlfriend-encouraged-suicide, https://futurism.com/character-ai-suicide-free-speech
[31] https://www.rollingstone.com/culture/culture-features/chatgpt-obsession-mental-breaktown-alex-taylor-suicide-1235368941/
[32] https://openai.com/index/how-we're-optimizing-chatgpt/

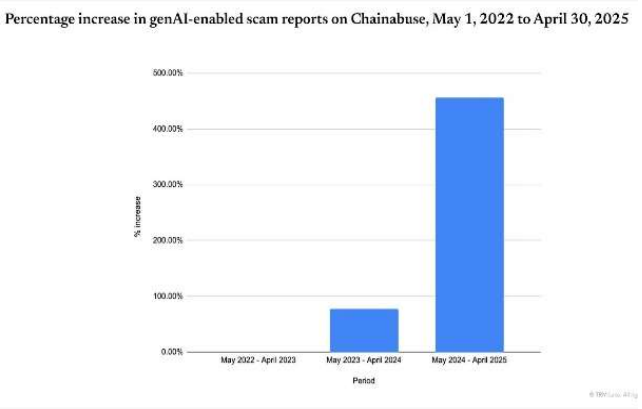
The cases show that mental problems can arise or worsen as a result of contact with language models such as ChatGPT, which can increase the challenges for aid workers such as the police.

Criminals tool

There are more and more malicious applications of synthetic media. Media that have been deliberately generated or manipulated to be used as a means in the context of crime or disinformation. The most common use still being deepporn, pornographic images with the faces of people who have not consented to it.[33] In the Netherlands, we are also witnessing a rapid increase in AI child pornography.[34]

In addition, we see GenAI being used extensively in social engineering, where communication is generated by language models, as well as the account information that, together with a deepfake profile photo, contributes to more authenticity. On LinkedIn, for example, this process is fully automated. Only when the intended victim has responded for the second time in a conversation, the conversation is forwarded to a real person. Emerging criminal phenomena in which GenAI plays an important role are fake employees who apply for remote work jobs[35] and dating fraud, for example aimed at motivating people to invest in crypto, via a link that refers to a rogue app or website.

In a recent report, TRM, the company behind the Chainabuse platform, shared the following explanation of GenAI's working method and role:[36]



Realistic Example:

A romance scammer runs a 10-step n8n workflow:

1. Scrapes dating profiles.
2. Scores emotional vulnerability with AI.
3. Sends intro messages using a GPT-powered agent.
4. Logs replies and engagement score.
5. If high, passes to a romance chatbot that continues relationship-building.
6. Prompts a move to a fake crypto platform.
7. Uses AI-generated dashboards to simulate profits.
8. When victim requests withdrawal, escalates to a "support agent" (another AI).
9. Requests tax payment or gas fees.
10. If victim pays, logs success and archives the case.

Other common uses are:

- *Scam & (identity) fraud*
  This involves using non-existent characters or cloning the voice and/or face of a trusted person for friend-in-need- imposter, CEO and emergency scams.

[33] https://offlimits.nl/assets/downloadable_files/rapport-onderzoek-toezicht-en-handhaving-online-seksueel-geweld.pdf
[34] https://www.nrc.nl/nieuws/2025/03/07/dark-llms-en-deepfakes-de-opmars-van-ai-kinderporno-a4885608 https://www.ad.nl/domestic/major blow-Europol-against-ai-child pornography-25-arrests-four-Dutch-buyers-suspicious~a7d0fdf5/
[35] https://www.linkedin.com/pulse/unmasking-remote-fake-workers-ismael-alvarez-rkqxe gives a good insight into how this works and what the motives are. Geopolitical motives can also play a role in this form of crime.
[36] https://www.trmlabs.com/resources/blog/ai-enabled-fraud-how-scammers-are-exploiting-generative-ai

Criminals make use of our ingrained tendency to believe what we see/hear ourselves. All of our evolution we have learned to trust in our own perception. Our brains are not used to the digital filter. So as soon as you see in your phone screen that the person you know well is calling (because the phone number is also spoofed/faked), your brain immediately adjusts to that person. The voice clone doesn't even have to sound very good anymore because your brain automatically dismisses any imperfections as a poor connection or environmental noise. There is no room for doubt even when the phone number is not spoofed, the familiar timbre of the well-known voice is enough to fool your brain.
Criminals can also use this to present themselves to a citizen by telephone as the trusted (neighbourhood) cop who then sends 'colleagues', or as a well-known 'colleague' who calls the department to obtain confidential information.

- *Deception online biometric authentication systems.*

  The quality of both face and voice clones is now so good that they can fool biometric authentication systems. This affected the Australian tax authorities[37], where you could log in based on voice verification, but also online systems for opening a bank or crypto account[3839].

All information has been manipulated

There is a lot of attention for rogue deepfakes and rightly so because it is becoming easier to clone a face or voice. But if you look more broadly, you see that deepfakes are part of the larger trend where everything becomes synthetic (fake).[40] This trend is much more important when it comes to disrupting processes, it affects our fundamental assumptions.

AI-generated or manipulated media are usually applied for entertainment, services or to improve our quality of life. But as police, we almost always use this information in a different context than what it was originally generated for. Additions to photos by AI in social media apps make it difficult to find out where a photo was taken. The fact that AI is embedded in more and more systems and that digital content will soon be manipulated by default makes police work much more complex.

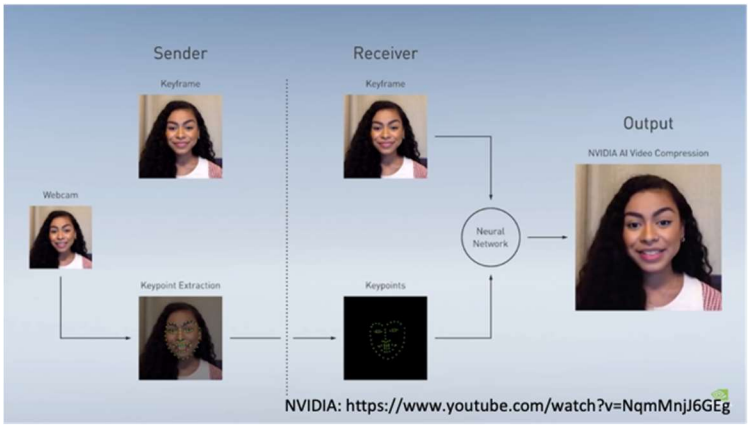
A good example to illustrate this is video calling, where GenAI has also made its appearance[41]. For example by taking a snapshot of your face at the beginning of the conversation to save bandwidt, after which, during the call, only a few points cross the line. With the snapshot and these points your face is re-generated by artificial intelligence on the receiving side.

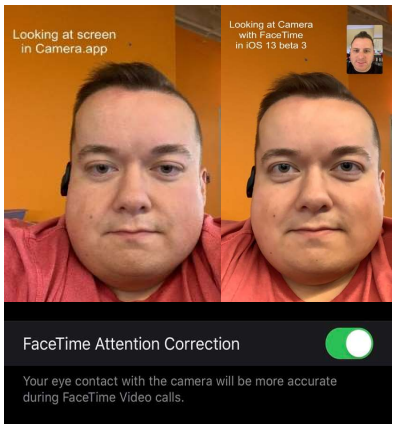
In fact, a number of applications now allows you to make you to look each other in the eye in a natural way, while you might actually be engaged in other things[4243]. All fake, but purely to improve the service, because it's about creating that familiar, intimate feeling, as if you were

[37]https://www.theguardian.com/technology/2023/mar/16/voice-system-used-to-verify-identity-by-centrelink-can-be-fooled-by-ai
[38] https://www.404media.co/inside-the-underground-site-where-ai-neural-networks-churns-out-fake-ids-onlyfake/
[39] Because more and more banks in the Netherlands also use the NFC chip in their identity documents when onboarding, it is hardly possible to open a regular bank account. However, online crypto and international bank accounts can be opened.
[40] https://www.bloomsbury.com/us/regulating-the-synthetic-society-9781509974948/, See also Deepfakes: The Coming Infocalypse by Nina Schick (2020)
[41] https://www.youtube.com/watch?v=NqmMnjJ6GEg
[42]Facetime image comes from a now deleted account on X: https://twitter.com/WSig/status/1146146914985009154.

[43]https://appleinsider.com/articles/20/06/22/facetime-eye-contact-correction-feature-to-launch-with-ios-14 in a small sample, this function was found to be enabled by default in the iPhones.

actually physically sitting across from each other and looking at each other during the conversation.





This simple example clearly shows how our perception is already affected and the importance of context. The manipulation in a videocall is not relevant in the context of a meeting or webinar. But we have to understand it when we are talking to a victim, witness or someone filing a report, or as a doctor with a patient. Unconsciously, we might also pay attention to non-verbal communication to assess trustworthiness.

But what is still reliable if only a few points of the face cross the line? After which, based on the averages of a very large group of people, your face is generated on the receiving side. What applies to the masses does not have to apply to you as an individual.

Before we can draw conclusions about digital content or communication via the digital channels, it is important to understand and assess the reliability and usability in our specific context.

Deployment of capacity, who is in charge

In addition to using information for evidence, our guidance is also based on it. Information determines how and where we deploy our resources. Previously, reference was made to the unnecessary use of resources to trace non-existing persons. This also applies to child pornography, where the police wants to identify the child as soon as possible. Due to the increasing quality of GenAI-generated images, it will become increasingly difficult to determine which child really exists.

GenAI can be used to disrupt public order, for example by having public figures or reliable sources say and do things they never did.[44] It can also be used to disrupt the chain-of-command during a crisis.

Another example is making false reports about non-existent emergencies, including supporting images. In America they now suffer from the Swatting service[45] where GenAI is used to direct the SWAT unit. This service is used by young people who are afraid of not passing a test. They register the name of their school on the website, pay $75 and the next morning the local police

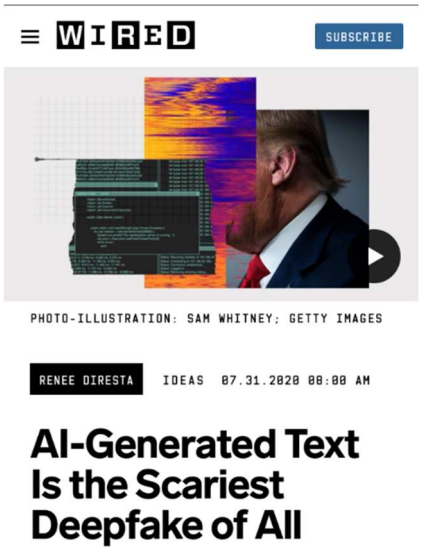
[44] In Baltimore, an audio recording was distributed in which a headmaster was racist. This deepfake caused great unrest. https://www.bbc.com/news/world-us-canada-68907895.
[45] https://www.vice.com/en/article/torswats-computer-generated-ai-voice-swatting/

are called by terrified (generated, synthetic) voices of non-existing children that are supposedly hidden in the school because someone with a gun walks through the hallway.

AI is increasingly being used to regularly update accounts or channels on social media with new material because of the advertising revenue. A fully AI-generated True Crime channel fed conspiracy theories about the failure of the mainstream media, as they paid no attention to these violent cases. The police were also addressed for their lack of commitment [46] and it took a while before they understood that these were made-up cases.

The internet is flooded with AI-generated content[47], not only on social media, but also sites like Amazon suffer from generated ads and generated books supposedly published by well-known writers or about popular topics. This can also be dangerous, as with books about picking mushrooms yourself, which contain recipes with poisonous mushrooms[48].

Social media platforms have an interest in this content, because they want to keep people engaged within their platform. In order to generate sufficient advertising revenue, it is necessary that there is always new content. For people it is difficult to post new content on a daily basis, but with generative AI you can produce and post dozens of videos per day. In four of the top 10 most popular channels on Youtube each video contains GenAI content.[49] The views and comments under content are also increasingly generated with AI. This makes it more difficult for the police to rely on social media for evidence or deployment of resources.



The exponential growth of this generated (nonsense) content, also known as AI Slop, leads to visible disruption of our global information ecosystem.[50] This ecosystem depends on the internet. That's why Wired already pointed to the danger of AI-generated texts in 2020.[51] A time when it was not even clear that within 5 years, anyone would be able to use these tools to generate high-quality content on any subject, in any language, tailored to any target audience..

Because of all this AI Slop, mis- and disinformation, it is becoming increasingly difficult to find reliable and relevant information, just as it is becoming increasingly difficult to get your information under the attention of others.

Disinformation

Disinformation is the deliberate dissemination of misleading information. The functioning of our brain in combination with the corresponding functioning of many well-known social media platforms and search engines facilitate both the rapid spread and the impact of disinformation.

[46] https://www.404media.co/a-true-crime-documentary-series-has-millions-of-views-the-murders-are-all-ai-generated/
[47] https://www.404media.co/ai-slop-is-a-brute-force-attack-on-the-algorithms-that-control-reality/
[48] https://www.404media.co/ai-generated-mushroom-foraging-books-amazon/
[49] https://sherwood.news/tech/ai-created-videos-are-quietly-taking-over-youtube/
[50] See also Deepfakes: The Coming Infocalypse by Nina Schick (2020)
[51] https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all/

Disinformation is often presented in a way that it evokes emotions and the urge to do something, even if it is just a like, share or comment. This so-called engagement is an important measure for prioritizing a message in news feeds, timelines and search engines.

By not only outsourcing production and distribution to AI, but also using AI to respond to your content, you increase engagement and which results in your (dis)information getting higher in the timelines. In addition, the following impacts are used in the effective dissemination of disinformation:

- Majority Illusion
  The underlying structure of social networks can ensure that a behavior or opinion that is rare worldwide can be systematically overrepresented in a local environment, because certain key figures share it. This is called a 'majority illusion'. It seems that everyone shares a certain opinion, which makes you inclined to go along with it. This is also used by so-called 'trollbots' that add automated fake opinions to discussions in large numbers. You see that more and more deepfakes are being used, both to make the accounts and the messages more authentic;
- Illusion of truth
  This concept concerns our tendency to see false information as real if we are repeatedly exposed to it. This has to do with our brain trusting things we hear and see for ourselves. Even though we know that something is not right, we (unconsciously) still believe in it because we encounter it so often. Which is why, even when shown that an image is fake, some people comment that the image might be fake, but the message behind it is not.[52] This makes the undermining effect of synthetic media extra large. What we hear and see is so authentic that it becomes difficult to continue to believe that it is fake.

Realistic audiovisual media are becoming increasingly easy to generate with AI. Due to the increasing quality, they not only provide extra authenticity, but can also evoke more emotions. This means that even more people will respond to a message or pay attention to it. GenAI is therefore a strong facilitator, but fortunately disinformation is still rare in the Netherlands and the impact is still low[53].

4. Practical guidance for law enforcement

The above illustrates the challenges law enforcement faces. It is becoming increasingly difficult to determine whether information is reliable. Several response options media are discussed in this chapter.

Practical handles

Earlier in this article, a distinction was made between specific GenAI, the traditional deepfakes, and generic GenAI. The reason for this is that the current deepfake detection tools focus on traditional deepfakes, they are specifically trained to do so. This is why images manipulated with other tools, despite being clearly fake to us, are not recognized as such by the deepfake detection tools.

[52] AI Slop: Last Week Tonight with John Oliver (HBO), https://www.youtube.com/watch?v=TWpg1RmzAbc
[53] https://decorrespondent.nl/15411/why we should have less-about-disinformation. What is growing strongly are AI-generated videos that, for example, glorify hitler and money

Especially when it comes to generic GenAI, there are still practical options available to everyone. Developments are fast[54], so the options mentioned are indications, not evidence! In addition, they relate to the pure results of GenAI, where no post-processing has taken place!

GenAI in general:

- **1 = None**. Generative AI always creates (generates) images from scratch, so that in different images the manipulations look different. Ask, for example a report is filed, for more photos or videos, preferably taken from different angles.
- Is it an outdoor environment? Compare with the real situation via Google Streetview.
- Look at the properties (exif/metadata) of a file. Secure the images yourselve from the source or device. The more likely the metadata is still available.
- See if you can find the photo on the internet by image searching. Sometimes you will find the same photo without manipulation, the date and context of that photo can give clues.
- AI is (for the time being) bad at the laws of physics and therefore less good in shadows or reflections in windows or pools of water. Often details in the background are not accurate.
  The fake picture on the right of Trump having breakfast with the Dutch Royals clearly shows such details.
  Trump's left hand is too small, the knife lies unnaturally under the coffee cup and the deformation of the spoon in the glass is not correct.
  But also from the context it can immediately be concluded that the image must be fake. At a royal breakfast you don't sit so close together, and an egg doesn't belong in the fruit bowl.

Deepfakes:

- Videos or photos in which someone does or says something that they actually haven't, are usually made with traditional deepfake tools. The current quality is so good that you no longer see imperfections in the images themselves. Technical and context-oriented research is therefor necessary, such as searching if the same images occur elsewhere, whether there is alternative information (Osint), checking if shadows & (weather) conditions match with claimed date/time, looking at metadata and asking domain experts or potential witnesses for more information.
- In live video connections ask the person to take his camera (laptop/phone) in his hand and turn around in the room while moving. GenAI is not yet able to compensate well for situations in which both the camera and the environment / person moves. It is of course better that you initiate the contact yourself via the telephone number or email address known to you.

[54] An up-to-date overview is maintained at https://digitalks.eu/tips-for-recognizing-of-genai

Deepfake detection tools, only for specialists as part of a wider toolbox

Together with the University of Amsterdam, the Netherlands Forensic Institute (NFI) has been researching deepfake detection tools for years.[55] The outcome is that these tools are unreliable.[56]

In May 2025, the NFI presented their research on blood flow detection as a possible future addition to the specialist toolbox to detect deepfakes. This method still needs to be validated.[57] Hopefully this will come in time, as research has shown that recent deepfakes are taking over the heart rate pattern from the original material. This eliminates the method based on the subtle volume changes of a blood vessel caused by the heartbeat.[58]

Even if they were good, detection tools have limited added value. They are trained on specific GenAI while the increase in synthetic media comes mainly from generic GenAI. And if they were to work for that, they would almost continuously have to warn us now that GenAI is being applied everywhere. The distinction between fake and real therefore becomes irrelevant. In fact, the term 'fake' no longer has any value; what matters is the reliability and applicability of the information in question in a specific context. Think of the AI filters on social media like Snapchat, in most cases these are not relevant. But perhaps certain filters can lead to a person not being recognized when you use the image for facial comparison.

Information manipulated with AI can also be reliable and relevant. To determine this, it is important to understand what manipulations have taken place. This is where a combination of technical tools can help. The NFI works according to this methodology, they use a diverse combination of, sometimes very specialist, technical tools, each with their own uncertainties, to arrive at a statement about reliability. It still requires a considerable investment in knowledge and expertise[59] before the police can embed this approach internally. And of course this also applies to the Public Prosecution Service and the Judiciary.

Regulation

According to the European AI Act, AI-generated content[60] must be labelled as such by the platforms. In addition, developers of (tools for) synthetic media must ensure that these can be detected, for example by watermarks.

It is already clear that the platforms do not meet the labeling requirement. For example, the aforementioned popular GenAI channels on Youtube do not always mention that AI is used. Also, the internet is full of tutorials on how to remove watermarks from GenAI content.

---

[55] https://www.forensischinstituut.nl/current/news/2022/11/15/new-methods-nfi-en-uva-for-recognize-deepfakes The manual method discussed for visual inspection yields less and less because the deepfaketools have been greatly improved. In the meantime, we are looking at which tools can be of added value per case. In practice, it remains a big challenge, for example if there is only a screenshot of an image, so that you also have no metadata.

[56] There are still regular studies that should show that a tool successfully recognizes more than 90% of deepfakes. In practice, these tools turn out to be tested on outdated datasets. Overall you can say that deepfakes made with outdated (often free) technology can indeed be recognized for more than 90% successfully. When testing for newer deepfakes (max 2 to 3 years old), this percentage drops to 50%. This means that on 100 images, 50 can be mistakenly referred to as fake. This makes use of these tools by the police irresponsible, they can not afford to put reliable information aside as a fake.

[57] https://www.forensischinstituut.nl/actueel/achtergrondverhalen-nfi/hoe-onze-hartslag-kan-verraden-dat-een-video-deepfake-is---eafs-2025-deel-2

[58] https://www.frontiersin.org/journals/imaging/articles/10.3389/fimag.2025.1504551/full

[59] This includes, for example, being able to deal with likelyhood ratios, which means if something is 60% likely. And how do you combine that with a 40% probability rate from another tool. What does the stacking of assumptions mean etc.

[60] https://www.technologyreview.com/2024/03/19/1089919/the-ai-act-is-done-heres-what-will-and-wont-change/

The question is not whether the EU will enforce strict enough to ensure that the regulation is complied with. The real question is what the added value of such regulation will be if almost all content is generated or manipulated by AI in any way.

Here, too, the distinction between AI-manipulated and non-AI-manipulated becomes irrelevant if the use is so widespread and GenAI is integrated into all kinds of applications, such as our telephone cameras, by default. The real need lies in being able to easily interpret the reliability and counteract malicious applications.

Provenance: grip on reliability for the general public

Not only within truth finding it is important to indicate the reliability of information, this also plays a major role in, for example, disinformation. Provenance goes beyond simple labeling or watermarking. The English concept of Provenance[61] can be translated as origin, but the meaning in this context is broader. This involves digitally irrefutable recording of proofs of origin, induced changes or the immutability of images and associated metadata using cryptographic encryption. It is therefore not just about recording and communicating the origin of information and the processes and methodology with which it was produced at the time of its creation. It is also about making it clear and being able to check all intermediate steps until the digital content ultimately reaches the final consumer of the content.

The beauty of provenance is on the one hand that you as a police or other organization can allow people to easily check whether the information really comes from you and has not been manipulated in the meantime. This builds trust in the police. On the other hand, you avoid discussions about what is or is not disinformation. People can always check whether the information really comes from the source concerned, but then have the freedom to decide for themselves whether they trust this source.

The international Content Authenticity Initiative (CAI)[62] to which parties such as Microsoft, Adobe, the BBC, Samsung, Canon and Reuters are affiliated, has already delivered various tools and standards for this.

Preventing Malicious Use

Enforcement and regulation are especially important when it comes to applications that do not deter or even encourage the use of GenAI in a rogue sense. Think of Grok, the LLM marketed by Elon Musk that has no limits, making it easy to create nude images.[63] But also the various deepporn and undressing (nudify) apps.

The current legislation already allows action to be taken against this. But such apps are still in the Appstore of, among others, Apple and are only be removed if journalists pay attention to

[61] On 26 January 2023, the first national working conference on the impact of synthetic media & deepfakes took place. The report and the green paper on provenance are published here: https://digitalks.eu/VerslagWerkconferentieImpactSynthetischeMedia

[62] https://contentauthenticity.org/

[63] https://www.theverge.com/report/718975/xai-grok-imagine-taylor-swifty-deepfake-nudes

them[64]. Active and coercive enforcement has a real effect as Apple[65] and France[66] have previously shown in their approach to Telegram.

Denmark has chosen to give people copyright on their own bodies, facial features and voice. This gives Danes the right to require platforms to remove material posted without permission.[67]

But, as argued by Etienne Valk of Utrecht University's[68] Institute for Information Law, the problem with copyright is that it is transferable. With an attractive offer from Big Tech or others, people can be tempted to transfer the rights so that they'll lose all control[69]. In the Netherlands expansion of the portrait right is a lot more logical because the extensive transfer of control over the portrait right is not possible. In addition, as Valk points out: *It is simply necessary not only to cover a person's face, but also their voice and body.* This broadening is really necessary because more and more (behavioural characteristics) are being measured, analysed and used, including in the context of identification.

In all cases, it remains important that the government takes enforcement action.

Gaining more certainty as a professional standard

For law enforcement, the biggest gains are in awareness, the ability to reflect critically and regular consultation with colleagues and experts inside and outside the police.

First of all, we need to be aware that many of our assumptions, based on years of experience, no longer apply. Such as the suicide case, where it was still assumed that a voice can only be used by the person herself. Furthermore, it is also important to understand that we ourselves can also be affected by disinformation. Our brains are also inclined to automatically trust that what we hear and see ourselves. We too can be called by someone who abuses the voice of a trusted colleague, or draw conclusions too quickly if we think we recognize a voice in an investigation. And finally, the awareness that even if there are no malicious intentions at play, we still have to look at manipulations, because they can be relevant in the context of our investigation.

Regularly this will give rise to additional research, in order to get more certainty. The aforementioned tools will help with this, as will the framework for action on how to deal with potentially AI-generated or manipulated voices[70].

The legal framework also takes into account that in the vast majority of cases for criminalisation it does not matter whether it is a deepfake, such as in the case of scams and extortion. It is mostly relevant when it comes to establishing identity; did that person log in to the tax service,

---

[64] https://www.404media.co/congress-pushes-apple-to-remove-deepfake-apps-after-404-media-investigation/

[65] https://www.theverge.com/2018/2/5/16974710/apple-telegram-ios-app-store-removal-explanation-child-pornography-distribution

[66] https://apnews.com/article/france-telegram-pavel-durov-judgment-6e213d227458f330ed16e7fe221a696c
https://www.wired.com/story/pavel-durov-judgment-telegram-content-moderation/

[67] https://nos.nl/nieuwsuur/artikel/2574932-in-de-strijd-tegen-deepfakes-krijgen-denen-copyright-op-eigen-gezicht-en-stem
https://www.theguardian.com/technology/2025/jun/27/deepfakes-denmark-copyright-law-artificial-intelligence

[68] https://www.nrc.nl/nieuws/2025/07/29/optreden-tegen-deepfakes-is-een-goed-idee-maar-doe-dat-dan-wel-via-het-portretrecht-a4901612

[69] https://www.businessinsider.com/background-actor-extra-body-scans-hollywood-ai-fears-report-2023-8
https://www.npr.org/2023/08/02/1190605685/movie-extras-worry-theyll-be-replaced-by-ai-hollywood-is-already-doing-body-scan People cannot oversee the consequences:
https://www.nytimes.com/2025/08/17/business/tiktok-ai-avatars.html?unlocked_article_code=1.fE8.hv1i.Tyof97BfHZ7X&smid=url-share.

[70] Internal publication Netherlands Police from February 2025.

was that suspect indeed involved in the intercepted conversation, or is this indeed the abducted victim.

Transparency of trade-offs made

In order to make good quality decisions, it is important to have sufficient certainty about the reliability of the information that is being worked with. The greater the impact of the decision to be taken, the more certainty is needed. If images containing someone who commits a criminal offence are now submitted by citizens, you have to wonder whether the images have been manipulated. In the case of images that neighbors have made of each other, you can ask the local community officer for more context. And if you do not get extra certainty about the (un)reliability, it might be better to invite the person concerned to the station instead of kicking in the door at five in the morning.

Law enforcement continuously works with uncertainty, it is often difficult to determine what is really true, but you can consider a certain scenario more plausible than other scenarios. When it comes to camera images of a store, you can also wonder whether someone had the opportunity to manipulate the images and whether it is technically plausible that the images have been manipulated. Are there multiple cameras that have taken images of the same situation from different angles? How plausible is it that multiple camera images have been manipulated at the same time?

This reflection (considerations) and the related research steps must be transparent, on the one hand in order to remain testable and reliable as an institute, on the other hand for very practical reasons. In the courtroom, the defence can claim that evidence is faked.[71] By already looking at possible manipulations during the investigation and explaining in the report that this has been done and how it affected the evidential value, the quality of the investigation increases and it saves resources. Police officers do not have to re-read the case in order to answer questions from the judge sometimes weeks or months after submission.

5. Conclusion: investing in critical capacity, reflection and expertise

We are at a tipping point. GenAI has, almost unnoticed, entered our daily lives through our phones, online platforms, apps, camera and communication systems, smart home devices and vehicles.

While GenAI can make crime and disinformation more efficient and effective, the biggest impact for law enforcement, as an information-intensive organization, is the disruption of the information ecosystem.

Although seldom malicious, AI-generated manipulations do affect our perception. They are created within specific contexts, such as services and entertainment, which are often different from the context in which law enforcement wants to apply the information. This can affect the value of evidence and decision-making.

Therefore, it is essential to gain both more certainty about the reliability of information and to deal with the increasing uncertainty. Only then can the police ensure the quality of its decision-making in the context of truth-finding and efficient and effective use of resources.

[71] For example, cases in which suspects claim that they did not open a bank account themselves, their defence is that this must have been a deepfake. The advice now is to always request the original images from the bank and already mention in the Verbal Procedure that this investigation has been done and what came out of it about the reliability.

The advice to the police is to:

1. To encourage the government to prevent the fraudulent use of A.I.;
    a. A much stricter enforcement of law & regulations;
    b. Expand portrait right with all biometric (behavioral) characteristics.
2. Provide citizens with tools to check the reliability of (police) information;
   By integrating provenance into all government and police generated, produced and shared information.
3. Invest within the police organisation and its chain partners in:
    a. Wide awareness and knowledge of action perspective;
    b. Leadership that encourages critical reflection and questioning of assumptions by taking an exemplary role and visibly questioning colleagues;
    c. Supporting expertises;
    d. Keeping track of developments;
4. Encourage scientific institutions to further assess the social impact of the use and misuse of A.I., as well as technologies that contribute to ensuring the authenticity and reliability of information.