The Threats of Manipulative Information Technologies

by Oliver Reindl

Introduction

With the rapid advancement and growing availability of powerful AI models, not only are new opportunities emerging in business, research, and daily life – but also new and serious risks. Technologies once restricted to research labs are now accessible to virtually anyone with an internet connection: text and image generators, voice imitation tools, deepfake software, and automated recommendation systems.
Public debate around artificial intelligence is often dominated by apocalyptic scenarios – the fear of AI replacing jobs, becoming uncontrollable, or even threatening humanity itself. However, most leading experts emphasize that AI, by itself, does not take away jobs. Instead, jobs may be lost to people who use AI and become significantly more productive with it. This makes it all the more urgent for professionals and organizations to actively engage with AI technologies rather than ignore or fear them.
Moreover, scientific studies have shown that AI systems depend on real-world data from humans to learn and improve. Models trained exclusively on synthetic (machine-generated) data tend to degrade or collapse over time. This illustrates a fundamental truth: AI needs humans – not only as users, but also as the source of authentic input and evaluation.
Yet beyond these long-term risks and philosophical questions, AI already poses very concrete threats today. There is a growing number of documented cases where AI is misused to deceive, manipulate, or harm people – through fake news, synthetic media, personalized scams, or algorithmic exploitation of behavioral patterns. These developments are not hypothetical. They are real, widespread, and increasingly sophisticated.
This study aims to present and explain the key information technologies that are currently being used to manipulate or deceive people. Each section includes a plain-language explanation, a technical description for IT professionals, and a real-world example to demonstrate the potential impact.

1. Deepfakes – Fake videos of real people

- Simple Explanation:

An AI system analyzes many real videos or photos of a person. It learns how the face moves and what expressions are typical. Then, it overlays the face onto another video. The result: it looks like the real person said or did something that never actually happened.

- Technical Description:

Deepfakes are primarily based on Generative Adversarial Networks (GANs), where a generator network creates images or videos and a discriminator network tries to detect whether the content is fake. Encoder-decoder architectures are also used to extract facial expressions and apply them using motion transfer techniques for realistic mapping.

- Use Case (Example):

A video shows a well-known politician allegedly admitting to election fraud. The convincingly fake video spreads rapidly on social media, undermining trust in democratic processes.

2. Voice Cloning – Imitating voices using AI

- Simple Explanation:

With just a few seconds of recorded speech, an AI can imitate a person's voice. It analyzes pitch, tone, and rhythm, and can then read any text aloud in that person's voice.

- Technical Description:

Voice cloning is powered by advanced Text-to-Speech (TTS) systems such as Tacotron 2 or VALL-E. These models extract acoustic features using an encoder and generate a synthetic speech representation using an autoregressive decoder. A neural vocoder (e.g., WaveNet, HiFi-GAN) then produces the final audio signal.

- Use Case (Example):

A scammer calls a bank pretending to be a company's CEO, requesting an urgent transfer – using the real voice of the executive.

3. AI Image Generators – Pictures that never existed

- Simple Explanation:

You describe what you want to see in text, e.g., 'a politician holding a suitcase of money'. The AI creates a realistic image that looks like a photo but is entirely fictional.

- Technical Description:

Image generators like DALL·E or Stable Diffusion use Latent Diffusion Models. These systems generate an image from random noise, guided by neural networks trained on billions of text-image pairs. The model incrementally transforms the noise into a coherent image that matches the input text.

- Use Case (Example):

A fake image shows a politician allegedly accepting bribes. Although the image is completely fictional, it spreads on social media and many people believe it's real.

4. AI-generated Fake News / Text

- Simple Explanation:

Large language models can write articles, stories, or comments that sound like they were written by real people – even though they are made up.

- *Technical Description:*

Language models like GPT-4 are based on Transformer architecture with self-attention. They are trained on large text corpora and generate content by predicting the most likely next word based on the given context.

🔹 Use Case (Example):
A blog post pretends to be a scientific article with false claims about vaccines. It spreads in anti-vaccine groups and is taken as fact.

5. AI-Powered Phishing – Perfect scam emails

🔹 Simple Explanation:
AI-generated emails look convincing, use logos, realistic language, and even refer to real information from social networks to gain trust.

🔹 Technical Description:
Phishing emails are generated using Natural Language Generation (NLG) and enhanced with NLP for personalization. Public data is analyzed to craft emails tailored to the recipient's role, writing style, or behavior. AI mimics formal language to increase credibility.

🔹 Use Case (Example):
An employee receives a fake email from 'IT support' asking to reset a password – leading to a fake website that collects login data.

6. Fake Reviews & Social Media Bots

🔹 Simple Explanation:
Automated programs post fake reviews or comments and share content on social media as if they were real people.

🔹 Technical Description:
Bots are programmed using scripting languages like Python and interact with platforms via public APIs or automation frameworks such as Selenium. They may use LLMs to generate human-like posts or follow predetermined text patterns.

🔹 Use Case (Example):
A company floods review platforms with fake 5-star ratings to promote a product that performs poorly in reality.

7. AR/VR to distort reality

🔹 Simple Explanation:
Virtual Reality (VR) and Augmented Reality (AR) show things that don't exist in the real world – yet can appear highly believable and immersive.

🔹 Technical Description:
AR/VR systems use 3D engines like Unity or Unreal and rely on sensor fusion for spatial tracking. Computer vision and real-time rendering are used to embed or simulate digital elements in physical environments.

🔹 Use Case (Example):
A fake VR educational app shows students a simulated version of history containing misinformation, which they believe is accurate.

8. GPS Spoofing – Faking your location

- Simple Explanation:

A device pretends to be at a different place by sending fake GPS signals – and apps believe it.

- Technical Description:

GPS spoofing uses Software Defined Radios (SDRs) to broadcast false satellite data. Since GPS lacks authentication, mobile receivers accept these signals as valid, allowing precise location manipulation.

- Use Case (Example):

A delivery driver fakes their GPS position to claim work hours or mileage that never happened.

9. Manipulative Recommendation Algorithms

- Simple Explanation:

Online platforms suggest content you're likely to engage with – often showing extreme or biased content to keep your attention.

- Technical Description:

Recommendation engines use collaborative filtering, matrix factorization, or reinforcement learning to optimize for user engagement. This can reinforce echo chambers or steer users toward increasingly extreme content.

- Use Case (Example):

After watching a few economic videos, a user is flooded with conspiracy-laden content about financial collapse – shaping their worldview.

10. Microtargeting – Ads aimed at your weaknesses

- Simple Explanation:

Websites track your behavior and show highly personalized ads – often with emotional triggers designed to influence you.

- Technical Description:

Tracking cookies, device fingerprinting, and data brokers create detailed user profiles. Ad-tech platforms like Google Ads use machine learning to segment and target users with precision using psychographic data.

- Use Case (Example):

An undecided voter receives political ads tailored to their fears, pushing them toward a specific party without realizing the manipulation.

11. Facial Recognition & Identity Theft

- Simple Explanation:

AI analyzes your face and can recreate or steal your identity for online impersonation or fraud.

- Technical Description:

Facial recognition uses deep learning via convolutional neural networks (CNNs) to extract

facial landmarks and convert them into embeddings. These vectors are used for comparison or reconstruction in verification systems.

- Use Case (Example):

A criminal creates a fake LinkedIn profile using your face and name to defraud your contacts.

12. Synthetic Videos in Politics

- Simple Explanation:

A politician is shown in a video making offensive statements – but the video was completely fabricated using AI.

- Technical Description:

Text-to-video platforms (e.g., Synthesia, Runway) use GANs and motion transfer, guided by text prompt embeddings, to generate synthetic videos. Facial expressions and gestures are simulated using deep motion capture models.

- Use Case (Example):

A fabricated video of a political candidate making racist remarks circulates before an election and damages their public image.

13. Dark Patterns – Hidden tricks in design

- Simple Explanation:

Websites and apps use clever design to trick you into doing things you wouldn't normally do – like accidentally subscribing or staying longer.

- Technical Description:

Dark patterns are optimized using A/B testing, heatmaps, and conversion tracking. UI/UX designs exploit cognitive biases (e.g., default bias, loss aversion) to guide user decisions in deceptive ways.

- Use Case (Example):

A user tries to cancel a subscription but can't find the option – or clicks the wrong button and gets charged for another year.

14. The recognition of falsifications

Forgery can only be detected if the original encoding is available. But if the fake picture, video or any digital product is copied, it is very difficult, possibly impossible to state whether it is real or fake.

Conclusion

This study has examined a wide range of current information technologies that can be used to manipulate or deceive people – intentionally or unintentionally. With each technology, we have provided:

• a simple explanation for general audiences,
• a technical description for professionals, and
• a practical use case to demonstrate real-world relevance.

What becomes clear is that artificial intelligence and related technologies are not inherently malicious – but they can become dangerous tools when misused. The power of AI lies in its ability to scale human behavior: for better or for worse.
Deepfakes, voice cloning, synthetic images, AI-written disinformation, targeted advertising, and manipulative design patterns are no longer abstract threats. They are already shaping opinions, decisions, and social dynamics on a large scale – often without the affected individuals being aware of it.
This reality demands both technical literacy and ethical awareness. Organizations, educators, policymakers, and the public must understand not just what AI can do, but how and why it is used – and by whom.
Only through transparent, responsible, and informed engagement can we ensure that AI strengthens rather than undermines trust, truth, and human dignity in the digital age.

Köln, June 2025.