

CENTRE FOR
STRATEGIC COMMUNICATION
AND COUNTERING DISINFORMATION

DEEPFAKE

User Guide

CONTENT:

Content.....1

Brief history, evolution and legal framework.....2

What is AI-generated content?.....3

How to identify fake materials generated by artificial intelligence.....4

Recommendations for citizens and civil society.....6

Conclusions.....7

SUMMARY

As artificial intelligence (AI) technology continues to advance, so do the methods used to generate misleading or false information. AI-generated content, often called "deepfakes" or synthetic media, has become a tool for disinformation campaigns that can easily deceive the public. To protect yourself and contribute to the integrity of Moldova's information ecosystem, it is essential to learn how to distinguish potential AI-generated fake material from real content.

This material, provided by the Centre for Strategic Communication and Countering Disinformation, outlines some steps and actions to help you recognise AI-generated materials and discern them from authentic content.

LEGAL AND REGULATORY FRAMEWORK

In the Republic of Moldova, the use of deepfake technologies can be framed under several existing legal provisions. The creation or dissemination of deepfake materials that mislead by attributing a false identity to a person may constitute identity fraud (art. 177^1 Criminal Code) or identity theft (art. 260^7 Criminal Code). Also, the use of deepfake for defamation purposes, the spread of false information or incitement to hatred and discrimination may be sanctioned as slander (art. 70 Contravention Code) or incitement to discrimination (art. 701 Contravention Code).

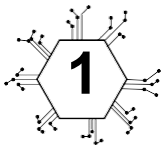
BRIEF HISTORY AND EVOLUTION

In a surprisingly short period of time, since its first appearances in 2017, deepfake technology has become extremely powerful, allowing for audio-visual manipulations of striking realism. Factors such as the development of artificial intelligence, increasing processing power and decreasing costs have made this technology not only accessible, but also a serious threat to cybersecurity and public trust. Now, when it is becoming increasingly difficult to distinguish what is real from what is fake, both organisations and each of us must act preventively to counter the associated risks and abuses.



WHAT IS AI-GENERATED CONTENT?

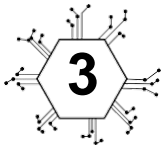
AI-generated content refers to text, images, audio, or video created or manipulated by artificial intelligence algorithms. This type of content is often designed to look and sound realistic, making it difficult to distinguish from authentic material. Examples include



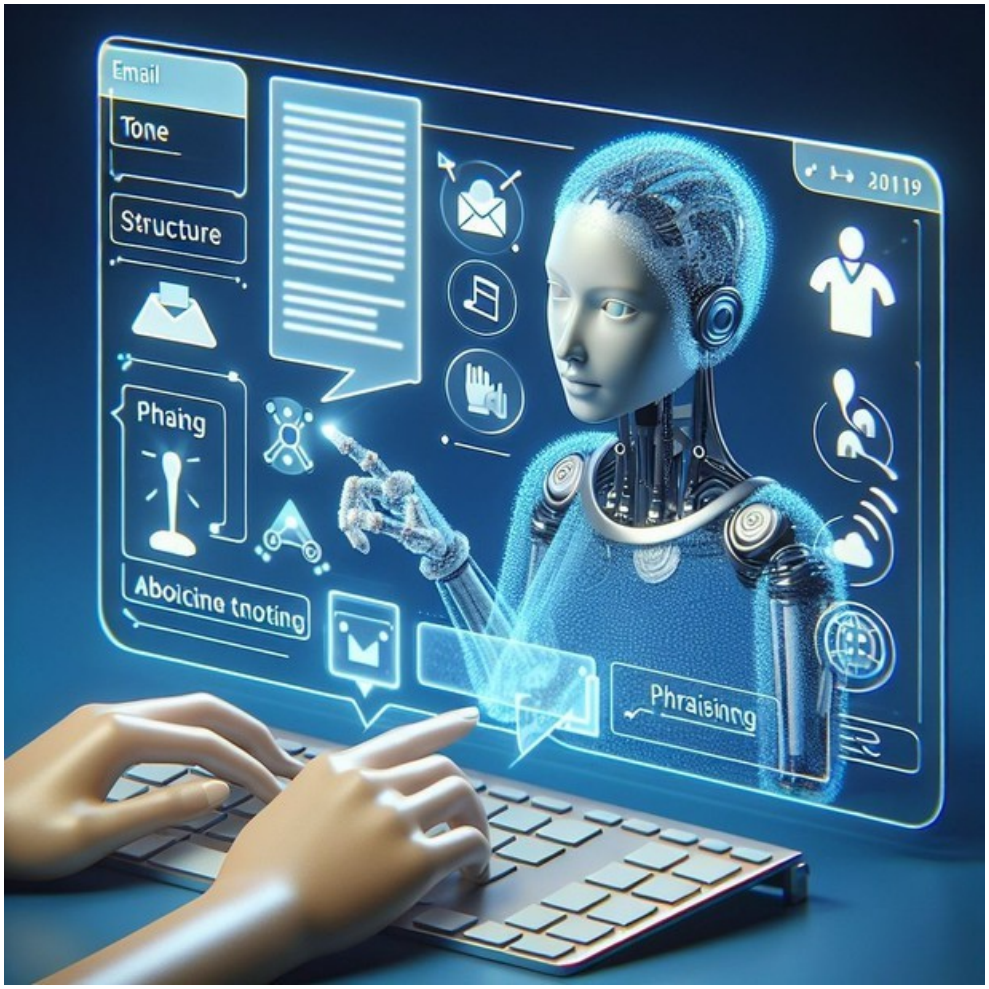
Deepfakes: videos showing people doing or saying things they have never done, generated using artificial intelligence.



AI-generated text: articles or social media posts that mimic human writing but contain misleading or incorrect information.



Synthetic audio: voice recordings generated by artificial intelligence that imitate real voices.





These forms of content can be used for disinformation, propaganda, or to damage the reputation of individuals and organisations.

HOW TO IDENTIFY FAKE MATERIALS GENERATED BY ARTIFICIAL INTELLIGENCE

Here are the specific steps and actions you can take to identify fake AI-generated content:

1. Examine the source of the content

- Check the source: Verify that the content comes from a reputable media outlet or official account. Be cautious with information from unknown websites or social media accounts.

- Verification symbol: Trusted sources often have verification symbols (blue checkmarks) on platforms like Facebook or Twitter. If the content comes from an unverified or suspicious source, be sceptical.

- Check with trusted media: If you encounter shocking or unusual content, check to see if trusted and established news sources are reporting the same information.

2. Analyse visual and audio quality

- Video anomalies: In deepfake videos, pay attention to details like inconsistent lighting, unnatural movements, or inappropriate facial expressions. Eyes may not blink naturally, and speech may not be perfectly synchronised with lip movements.

- Audio inconsistencies: In audio content, you may hear strange pauses, inappropriate tones, or unnatural inflections that don't sound like a real person.

- Image inconsistencies: AI-generated images may display irregularities such as blurry or pixelated areas, unrealistic reflections, or distorted objects, especially in the background or finer details (such as hands or eyes).



3. Check the content for logical defects

Weird statements: Be sceptical of sensational or outrageous claims that seem designed to provoke an emotional reaction. AI-generated text often exaggerates to manipulate emotions.

Inconsistent language: AI-generated text may lack coherence or suddenly change tone. Look for awkward phrases, repetitions, or grammatical errors that don't fit with the overall style.

Too good to be true: Some AI-generated images or videos are designed to look too perfect. If something looks too staged or polished, it could be synthetic media.



4. Use online tools to verify content

Deepfake detection tools: Use online platforms or browser extensions specifically designed to detect deepfakes or manipulated content, such as Deepware Scanner or Sensity.

Reverse image search: If you suspect an image may have been generated or modified by artificial intelligence, perform a reverse image search using platforms like Google or TinEye. This can help you find the original source of the image and see if it has been modified.

Check metadata: Use online tools to check the metadata of your videos or images. Metadata can provide clues about when, where, and how the content was created.

5. Be aware of context and time

Cross-reference with news events: Compare the content in question to real-world events. If the content doesn't align with timelines or confirmed events, it's probably fake.

Timing and Intent: Disinformation campaigns often release AI-generated content during sensitive times, such as elections, crises, or political unrest. Be very cautious with content during these times.

RECOMMENDATIONS FOR CITIZENS AND CIVIL SOCIETY

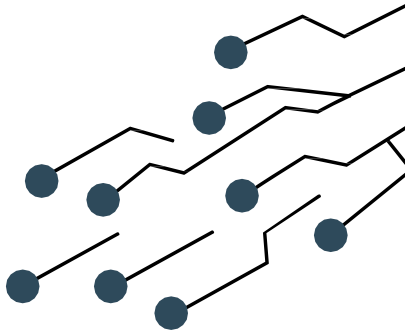
1. **Be critical and informed:** Develop the habit of questioning everything you see online. Don't share content before verifying its authenticity, especially if it evokes strong emotional responses.

2. **Educate others:** Spread awareness of AI-generated content and how to identify it among your friends, family, and community. Disinformation spreads easily, but so does truth when people are informed.

3. **Follow trusted sources:** Stay connected to trusted media outlets, government channels, and reputable organisations. Official channels will provide accurate updates during crises or events.

4. **Participate in media literacy programs:** Attend media literacy workshops or courses offered by the Centre for Strategic Communication and Countering Disinformation to improve your skills in identifying false content.

5. **Report suspicious content:** If you encounter suspicious materials generated by artificial intelligence, report them to fact-checking platforms or share your concerns with the competent institutions (the Centre) or media watchdogs. The Centre for Strategic Communication and Countering Disinformation offers citizens a platform to report disinformation campaigns and harmful content to the email addressdialog@stratcom.md.



CONCLUSION

In a world where AI-generated content is becoming increasingly prevalent, identifying fake material is crucial to protecting the information ecosystem and democracy in Moldova. By remaining vigilant, using the right tools, and sharing awareness, citizens and civil society can play a critical role in protecting truth and transparency in our digital world.

The Centre for Strategic Communication and Countering Disinformation is committed to helping you navigate these challenges and encourages all citizens of the Republic of Moldova to support the resilience of our information environment.

This article serves as a basis to increase awareness and the public's ability to detect AI-generated disinformation, contributing to protecting Moldova's national security and promoting an informed society.

