Information about **Deepfakes**

Autorité
luxembourgeoise
indépendante de
l'audiovisuel

Version: 12.01.2026

Autorité luxembourgeoise indépendante de l'audiovisuel (ALIA)

18, rue Erasme
L-1468 Luxembourg
+352 247-70 105
samra.cindrak@alia.etat.lu

What is a Deepfake?

Deepfakes refer to highly realistic content created or altered using artificial intelligence (AI). They can convincingly replicate a person's appearance, voice, or gestures making it seem as though they said or did something that never actually occurred. Beyond merely mimicking real individuals, such technologies are also capable of generating entirely fictional scenes or events that appear increasingly authentic. Deepfakes are a form of synthetic content, most often created with the intent to deceive, mislead, and cause harm.

Thanks to artificial intelligence (AI), it has now become significantly easier, faster, and cheaper to manipulate or create large amounts of images, videos, and audio recordings. Applications with intuitive interfaces, often built on open-source software, now enable anyone to create convincing deepfakes from a simple text prompt, without requiring advanced technical skills. By democratizing access to such technologies, these developments have expanded the number of people able to use them, thereby increasing the risk of malicious use and amplifying the potential impact of their harmful effects.

Despite their creative potential, these technologies can be exploited by malicious actors to:

Commit fraud

- Through various **spoofing** techniques, individuals can impersonate another person or, for example, a company in order to gain their victims' trust and deceive, manipulate, or defraud them. Deepfake-enhanced scams such as **CEO fraud** (or "president fraud"), in which criminals impersonate company executives to trick employees into revealing sensitive data or transferring funds, have risen sharply. For instance, in early 2024, an employee in Hong Kong transferred €23 million to scammers who simulated an entire videoconference using AI, convincing him he was speaking with his manager and the company's board of directors.
- **By bypassing biometric verification systems** (such as facial or voice recognition) **using deepfakes**, enabling fraudsters to gain access to smartphones, computers, applications, bank accounts and, consequently, sensitive data.
- **Through the use of deepfakes to enhance traditional fraud schemes**, such as the so-called "grandparent scam," "shock calls", "romance or love scams", or other forms of online fraud. Data on potential victims can be easily compiled from publicly available profiles and analysed to create detailed personal profiles or customised conversation scripts. Messages can be translated without error and disseminated on a large scale. In addition, the ability to automatically generate personalised responses significantly increases the scalability of fraudulent activities, enabling perpetrators to conduct numerous conversations in parallel and thereby reach a much larger number of people. Improved deepfake quality and increased message personalisation enhance the credibility of these fraud attempts. Many victims feel reassured when they believe they are speaking in real time with a familiar person whose voice or face they recognise during a phone or video call, without realising that it is an imitation. Once trust has been established, victims often disclose further sensitive information. Fraudsters can then build emotional ties and exploit these relationships to commit additional crimes, such

as grooming (the online approach of a minor for the purpose of sexual exploitation), cyberbullying, or “sextortion” (blackmail involving intimate images).

- In this context, **social engineering techniques** exploit cognitive biases and emotions to make deceptive narratives seem credible. Applied to deepfakes, they reinforce and accelerate all forms of scams by:
    - creating a sense of urgency or fear that short-circuits critical thinking;
    - impersonating the appearance, voice, or authority of trusted individuals to elicit obedience;
    - exploiting victims’ goodwill or empathy to extract information or money.

Discredit individuals or political opponents

Deepfakes can serve dual purposes: on one hand, to **discredit** or **defame** someone by attributing false statements or actions to them; on the other, to exploit the **credibility** and public image of well-known figures to lend weight and legitimacy to misleading messages.

Political figures are often the primary targets of manipulation campaigns. However, other public figures, such as well-known actors, prominent presenters or popular media brands, are also affected due to their high visibility and the trust they inspire. The large amount of publicly available audiovisual material about them further facilitates the use of this data to train AI systems, enabling the creation of highly realistic deepfakes.

Spread disinformation and manipulate public opinion

Deepfakes are increasingly used as part of coordinated disinformation campaigns designed to:

- **manipulate** public opinion;
- sow **doubt** and **undermine trust** in public institutions, the media, and democracy as a whole;
- compromise the integrity of **elections** and democratic processes;
- **divide society**.

AI-generated content can serve multiple manipulative purposes, particularly in the run-up to elections, posing a serious threat to democratic integrity.

Two days before the parliamentary elections in Slovakia in September 2023, for example, an AI-generated audio recording circulated in which opposition leader Michal Šimečka appeared to admit to manipulating the vote in his party’s favor. Shortly before the New Hampshire primary in the United States in January 2024, an automated robocall mimicking Joe Biden’s voice was sent out, apparently to discourage voters from going to the polls. As early as April 2023, the Republican Party had released, immediately after Biden announced his bid for re-election, an AI-generated campaign video depicting a dystopian scenario of his second term: scenes of chaos, economic collapse, war, and uncontrolled migration, all designed to dissuade voters. The case of Rumeen Farhana, an opposition politician in Bangladesh, whose AI-generated images showing her in a bikini circulated in late 2023, illustrates that significant harm can occur even when the manipulation is obvious. Although the fake was quickly exposed, the images provoked public outrage in a predominantly Muslim country. And then there is Donald Trump, who regularly shares deepfakes featuring himself, including a recent one showing him dropping

excrement from a fighter jet onto protesters from the “No Kings” movement, mocking opponents of his government who were demonstrating against his authoritarian drift and his tightening grip on power.

Malicious actors use deepfakes to influence electoral choices in various ways: by damaging the integrity of political opponents, deterring certain voters from participating in the election, or deliberately spreading false electoral information. These technologies are also used to stage fear-inducing or threatening scenarios, transforming abstract ideas into vivid images that make the conveyed messages more credible, striking, and persuasive. More broadly, they are employed to steer public opinion, often by mobilizing outrage around sensitive topics such as migration or refugee reception, to support specific political agendas. At the same time, they serve to reinforce and amplify pro-government narratives, while diverting attention from issues critical of those in power and saturating the information space with irrelevant, misleading or false information.

The proliferation of disinformation and the ease with which AI-generated content can mimic reality are undermining a **fundamental human capacity**: to understand the world through their **senses**, and to rely on what they see and hear to make sense of reality. This blurring of the boundary between truth and falsehood fosters the feeling that nothing is true anymore and can be verified, thereby undermining collective trust. This situation is likely to be further exacerbated by what researchers refer to as the “**liar’s dividend**”, which enables political actors to challenge, or even deny the authenticity of facts that prove inconvenient or unfavorable to them by invoking the possibility of deepfake manipulation. Thus, deepfakes become tools of strategic denial, contributing to the erosion of trust in the media, institutions, and even the concept of truth itself.

In January 2023, **Luxembourg**’s main political parties (DP, LSAP, déi gréng, ADR, Déi Lénk, Piraten, Fokus, and Volt) signed an electoral agreement committing to conduct fair and factual campaigns for the 2024 European elections. In this agreement, they expressed their intention to use social media responsibly and to refrain from personal attacks, the spread of false information, defamatory campaigns, and the tampering with or manipulation of other parties’ campaign materials. Although the text does not explicitly mention a ban on deepfakes, it does include a pledge not to use bots or automated propaganda or manipulation campaigns. It can therefore reasonably be inferred that such AI-based technologies should not be employed for disinformation purposes.

Produce illegal content

This category notably includes **deepnudes**: **pornographic** content generated without the consent of the people depicted. The production, possession, or distribution of child-pornographic material, whether AI-generated or not, that depicts minors is strictly prohibited and punishable by law.

Deepfakes are rarely disseminated in isolation; they are often embedded in **hybrid campaigns** that combine various AI-based tactics to maximize their reach and impact.

In Luxembourg, deepfakes have circulated featuring, among others, Prime Minister Luc Frieden, the Mayor of Luxembourg City Lydie Polfer, the abdicating Grand Duke Henri, as well as former Prime Minister Jean-Claude Juncker. In these manipulated or AI-generated videos and articles, the public figures appeared to give financial advice and encourage the public to invest through

cryptocurrency platforms. Journalists from RTL, Mariette Zenners, Caroline Mart, and Lynn Cruchten, as well as actress Désirée Nosbusch, were also victims of deepfakes used to make these scams more convincing.

The fraudsters combined several methods:

- reproducing the design of reputable news websites (such as RTL Luxembourg or Tagesschau), copying logos, colors, and layouts;
- creating fake articles with sensational headlines promising “exclusive revelations”;
- promoting these fakes via fake social-media profiles and paid advertising services on Facebook and Instagram.

What does the law provide for?

Legislation imposes obligations at several levels:

- on the **developers** of generative AI models,
- on the **platforms** where this content is distributed, and
- on the **users (deployers)** of these tools, when the AI is not being used strictly in a private context.

Obligations for AI developers

AI developers must:

- ensure that every person interacting with an AI system can recognize that they are communicating with a machine and not a human being;
- guarantee that any output produced or modified by AI can be identified as such through machine-readable markings. Examples include invisible watermarks or metadata indicating the date of creation or modification, the model used, and other identifying details.

Preventive measures are required to reduce the production of harmful deepfakes, such as:

- implementing input and output filters to prevent the generation of damaging deepfakes (especially those targeting minors, deepnudes, etc.).
- conducting regular vulnerability assessments on generative AI systems, as well as legal-compliance checks on the content they produce.

Obligations for Platforms

Platforms must:

- establish clear rules governing the creation and sharing of synthetic content, consistent with their obligation to moderate illegal material;
- provide for sanctions (such as account suspension or deletion) against users who violate these rules;
- enable users to report deepfakes that may constitute illegal content.

Platforms are also encouraged to:

- provide tools allowing users to indicate whether content has been generated or altered by AI, for instance, through an “AI notice” or label displayed below the relevant post.

Obligations for Users

Users must:

- clearly indicate when they generate a deepfake and refrain from creating harmful or illegal deepfakes;
- report any suspicious or misleading content.

How to recognise a deepfake

As AI models continue to advance, they produce obvious errors less frequently, leaving no consistently reliable visual or audio cues by which deepfakes can be clearly identified as such. In many cases, detection relies on analyses using specialised software or detailed manual examinations, such as frame-by-frame analysis or slow-motion playback, which, at this stage, do not yet guarantee complete reliability.

Many people mistakenly believe that content is authentic simply because they do not detect any “typical AI error.”

Detection therefore often depends on **context**, **content**, and **logical inconsistencies**. AI can imitate voices and gestures, but still struggles to understand physical laws or causality. Thus, illogical behavior in a video, for example, a person remaining motionless during a fire, or a car driving the wrong way down a bike lane, can break the illusion and reveal a deepfake.

What can I do?

- **Stay informed:** Read articles and subscribe to newsletters or podcasts about artificial intelligence. Technology evolves quickly: the better you understand it, the easier it becomes to spot fakes.
- **Check for AI labels**: The way AI labels are designed and displayed (color, text, symbol, placement) may vary depending on the platform and its interface. Often, the mention that an image or video is AI-generated appears in the caption. Look for labels or hashtags such as #AI, #GeneratedWithAI, or #deepfake, or for the logo of the tool used (for example, Sora by OpenAI).
- **Avoid sharing lightly**: Every like, comment, or share contributes to the spread of false content. If you have doubts about authenticity, do not share the material.
- **Verify media coverage**: Reputable media outlets report on important events quickly. If you cannot find any trace of a story elsewhere, it is probably false. Always check information by consulting several sources rather than relying on just one. You can also visit fact-checking websites such as Fact-checks – EDMO Belux or go RTL Today – Fact Check.

- **Approach content critically**: It is estimated that by 2026, nearly 90 % of online content will be generated by AI. Stay cautious, but avoid falling into a generalized distrust of legitimate journalistic or official sources. In a world where the line between reality and fiction is increasingly blurred, trust in transparent, responsible, and reliable communication is essential.

Links and useful resources

- Find essential information and practical advice to protect yourself against financial fraud at letzfin.lu, the reference platform for financial education run by Luxembourg's Commission de Surveillance du Secteur Financier (CSSF). The section "Précautions à prendre" (only available in French and German) raises awareness of the most common types of financial scams, including AI-assisted phishing, identity theft, and fake websites. Follow @letzfin on Instagram to stay up to date.
- The National Cybersecurity Competence Center (NC3), hosted at the Luxembourg House of Cybersecurity (LHC), provides companies and institutions with advice and practical tools to strengthen their digital security, including against attacks exploiting artificial intelligence.
- In cases of online fraud, the website cyberfraud.lu guides you toward the appropriate steps and competent authorities. Launched in June 2025 by the Luxembourg House of Cybersecurity (LHC) and the Association des Banques et Banquiers Luxembourg (ABBL) as part of a national awareness campaign, this initiative is supported by more than fifteen partners.
- For any general question about cybersecurity or online safety, contact the BEE SECURE Helpline at +352 8002 1234 or via the online contact form (only available in French and German).