Informationen zu **Deepfakes**

Autorité
luxembourgeoise
indépendante de
l'audiovisuel

Version vom 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

Was ist ein Deepfake?

Deepfakes sind Inhalte, die mithilfe künstlicher Intelligenz (KI) erstellt oder verändert wurden und täuschend echt wirken. Sie können das Aussehen, die Stimme oder die Gestik einer Person nachahmen und diese so scheinbar Dinge sagen oder tun lassen, die sie in Wirklichkeit nie gesagt oder getan hat. Über die Nachahmung realer Personen hinaus ermöglichen diese Technologien auch die zunehmend realistische Darstellung frei erfundener Szenen oder Ereignisse. Deepfakes werden daher auch als synthetische Inhalte bezeichnet, die in der Regel mit einer Täuschungsabsicht erstellt werden, etwa um irrezuführen, Meinungen gezielt zu beeinflussen oder Schaden zu verursachen.

Dank künstlicher Intelligenz ist es heute deutlich einfacher, schneller und kostengünstiger geworden, große Mengen Bild-, Video- und Audiomaterial zu manipulieren oder herzustellen. Anwendungen mit intuitiven, bedienungsfreundlichen Benutzeroberflächen, die häufig auf Open-Source-Software basieren, ermöglichen es inzwischen immer mehr Menschen ohne besondere technische Kenntnisse, mithilfe eines einfachen Textbefehls („prompt“) gefälschte Videos zu erstellen. Der erleichterte Zugang zu diesen Technologien hat den Nutzerkreis erheblich erweitert und damit zugleich das Risiko einer missbräuchlichen Anwendung sowie die Tragweite möglicher schädlicher Auswirkungen vergrößert.

Trotz ihres kreativen Potenzials können diese Technologien von böswilligen Akteuren missbräuchlich eingesetzt werden zu(r):

Betrugs- und Täuschungszwecken

- durch verschiedene Formen des sogenannten „**Spoofings**“, bei denen sich eine Person als jemand anderes, beispielsweise ein Unternehmen ausgibt, um das Vertrauen der Opfer zu gewinnen, diese zu täuschen, zu manipulieren oder zu betrügen. Insbesondere sogenannte „**CEO-Scams**“ (auch „Chef- oder Präsidentenbetrug“), die durch den Einsatz von Deepfakes immer überzeugender werden, nehmen deutlich zu. Dabei imitieren Kriminelle Führungskräfte eines Unternehmens, um Mitarbeitende zur Preisgabe sensibler Informationen oder zur Überweisung von Geldbeträgen zu bewegen. So überwies Anfang 2024 ein Mitarbeitender in Hongkong rund 23 Millionen Euro an Betrüger, die mithilfe künstlicher Intelligenz eine vollständige Videokonferenz simuliert hatten und ihn glauben ließen, mit seinem Vorgesetzten und dem Verwaltungsrat des Unternehmens zu sprechen.

- durch die Aushebelung **biometrischer Verifizierungssysteme** mithilfe von Deepfakes (z.B. Gesichts- oder Spracherkennung), die es Betrügern ermöglicht, sich Zugang zu Smartphones, Computern, Apps und Bankkonten zu verschaffen und so an sensible Daten zu gelangen.

- **Durch den Einsatz von KI zur Perfektionierung klassischer Betrugsmaschen** wie des sogenannten „Enkeltricks“, von Schockanrufen, Liebesbetrug („love scams“) oder anderen Formen des Online-Betrugs. Mithilfe von KI lassen sich Daten potenzieller Opfer leicht aus öffentlich zugänglichen Profilen zusammentragen und analysieren, um detaillierte Persönlichkeitsprofile oder individuell zugeschnittene Gesprächsskripte zu erstellen. Nachrichten können fehlerfrei übersetzt und massenhaft versendet werden. Zudem erhöht die Möglichkeit, automatisiert personalisierte Antworten zu generieren

die Skalierbarkeit betrügerischer Aktivitäten erheblich, da Täter zahlreiche Gespräche parallel laufen lassen und so eine deutlich größere Zahl von Personen erreichen können. Mit steigender Qualität und zunehmender Personalisierung von Deepfakes erscheinen diese Betrugsversuche immer überzeugender. Viele Betroffene wähnen sich in falscher Sicherheit, wenn sie davon ausgehen, in Echtzeit mit einer vertrauten Person zu sprechen, deren Stimme oder Gesicht sie in einem Telefon- oder Videoanruf wiederzuerkennen glauben, ohne zu ahnen, dass es sich um eine Täuschung handelt. Fassen die Betroffenen einmal Vertrauen, geben sie häufig weitere sensible Informationen preis. Auf dieser Grundlage können Täter emotionale Bindungen herstellen und gezielt ausnutzen, um weitere Straftaten, wie etwa „**Grooming**“ (die gezielte Online-Ansprache Minderjähriger zu sexuellen Zwecken), **Cybermobbing** oder „**Sextortion**“ (Erpressung mithilfe intimer Bilder) zu begehen.

- Hierbei kommen Techniken des sogenannten „**Social Engineering**“ zum Einsatz, die gezielt kognitive Verzerrungen und emotionale Reaktionen ausnutzen, um gefälschte Inhalte überzeugender wirken zu lassen. Deepfakes verstärken sämtliche Formen von Betrug, indem sie:
    - ein Gefühl von Dringlichkeit oder Angst erzeugen, das die kritische Urteilsfähigkeit untergräbt;
    - das Aussehen, die Stimme oder die Autorität bekannter oder angesehener Personen missbrauchen, um Vertrauen und Gehorsam hervorzurufen;
    - die Hilfsbereitschaft oder Empathie der Opfer ausnutzen, um diese zur Preisgabe sensibler Informationen oder zur Überweisung von Geldbeträgen zu bewegen.

Diskreditierung und Kompromittierung von Privatpersonen oder politischen Gegnern

Deepfakes können auf zweierlei Weise eingesetzt werden: Zum einen, um Personen gezielt zu diskreditieren oder diffamieren, indem ihnen Aussagen oder Handlungen zugeschrieben werden, die sie nie getätigt haben. Zum anderen wird das öffentliche Ansehen bekannter oder renommierter Persönlichkeiten und das Vertrauen, das ihnen entgegengebracht wird, gezielt ausgenutzt, um trügerischen Botschaften mehr Relevanz und Legitimität zu verleihen. Solche Manipulationskampagnen richten sich in erster Linie gegen politische Akteure. Aber auch andere Personen des öffentlichen Lebens wie bekannte Schauspieler, Moderatorinnen oder beliebte Medienmarken geraten aufgrund ihrer großen Reichweite und des Vertrauens, das sie genießen, ins Visier. Da von diesen Personen viel öffentlich verfügbares Bild- und Tonmaterial existiert, können KI-Systeme vergleichsweise leicht mit entsprechenden Daten gespeist werden, um sehr realistisch wirkende Deepfakes zu erzeugen.

Verbreitung von Desinformation und Manipulation

Deepfakes werden auch dazu genutzt, koordinierte Desinformationskampagnen zu erstellen und zu verbreiten, die darauf abzielen,

- die **öffentliche Meinung** zu manipulieren;
- Zweifel zu säen und das Vertrauen in **öffentliche Institutionen**, die **Medien** und die **Demokratie** insgesamt zu untergraben;
- die Integrität von **Wahlen** und demokratischen Prozessen zu beeinträchtigen;
- die **Gesellschaft** zu spalten.

KI-erzeugte Inhalte können auf vielfältige Weise zu manipulativen Zwecken eingesetzt werden, und stellen insbesondere im Vorfeld von Wahlen ein ernstzunehmendes Risiko für demokratische Prozesse dar.

So kursierte zwei Tage vor den Parlamentswahlen in der Slowakei im September 2023 eine KI-erzeugte Audioaufnahme, in der der liberale Oppositionsführer Michal Šimečka scheinbar einräumte, versucht zu haben, die Wahl zugunsten seiner Partei zu beeinflussen. Kurz vor den Vorwahlen im US-Bundesstaat New Hampshire im Januar 2024 wurde zudem ein automatisierter Anruf („Robocall“) abgesetzt, der die Stimme von Joe Biden imitierte und offenbar darauf abzielte, Wählerinnen und Wähler von der Stimmabgabe abzuhalten. Bereits im April 2023 hatte die Republikanische Partei unmittelbar nach der Bekanntgabe, dass Biden für eine zweite Amtszeit kandidieren würde, ein KI-generiertes Video veröffentlicht, das seine Wiederwahl als fiktives Katastrophenszenario inszenierte und die Wählerschaft mit Bildern von Chaos, wirtschaftlichem Zusammenbruch, Krieg und unkontrollierter Migration abschrecken sollte. Das Beispiel der Ende 2023 verbreiteten KI-generierten Bilder, die die Oppositionspolitikerin Rumeen Farhana aus Bangladesch im Bikini zeigten, verdeutlicht, dass selbst klar erkennbare Manipulation erheblichen Schaden anrichten kann. Obwohl die Fälschung rasch als solche erkannt wurde, lösten die Bilder in dem mehrheitlich muslimischen Land starke öffentliche Empörung aus. Und dann wäre da noch Donald Trump, der regelmäßig Deepfakes verbreitet, in denen er sich selbst inszeniert. Dazu zählt auch ein kürzlich veröffentlichtes Deepfake, das ihn dabei zeigt, wie er aus einem Kampfjet Exkremente auf Demonstrierende der „No Kings“- Bewegung abwirft, um politische Gegner zu verhöhnen, die gegen seine autoritäre Politik und seinen wachsenden Machtanspruch protestierten.

Böswillige Akteure greifen auf Deepfakes zurück, um **Wahlentscheidungen auf unterschiedliche Weise zu beeinflussen**: indem sie die Integrität politischer Gegner untergraben, bestimmte Wählerinnen und Wähler von der Stimmabgabe abhalten oder gezielt falsche wahlbezogene Informationen verbreiten. Darüber hinaus werden diese Technologien gezielt eingesetzt, um abstrakte Angst- oder Bedrohungsszenarien in konkrete Bilder zu übersetzen und die vermittelten Botschaften glaubwürdiger, eindringlicher und überzeugender erscheinen zu lassen. Neben der Manipulation von Wahlen können Deepfakes auch genutzt werden, um gezielt Einfluss auf die öffentliche Meinung zu nehmen und diese in bestimmte Richtungen zu lenken, um etwa gegen sensible Themen wie Migration oder die Aufnahme von Geflüchteten zu mobilisieren und so bestimmte politische Agenden zu stärken. Umgekehrt können sie auch dazu eingesetzt werden, regierungsfreundliche Narrative zu verbreiten, oder

aber von regierungskritischen oder unliebsamen Informationen abzulenken und den Informationsraum gezielt mit irrelevanten, irreführenden oder falschen Informationen zu überfluten.

Die zunehmende Verbreitung von Desinformation und die Möglichkeit, mithilfe künstlicher Intelligenz Inhalte zu erzeugen, die von der Realität kaum noch zu unterscheiden sind, untergraben eine **grundlegende menschliche Fähigkeit**: die Welt über die eigenen **Sinne zu erfassen** und sich auf das verlassen zu können, was sie sehen oder hören, um daraus Sinnzusammenhänge zu erschließen. Wenn die Grenze zwischen Wahrheit und Fälschung verschwimmt, entsteht das Gefühl, dass nichts mehr wahr oder überprüfbar ist, wodurch das kollektive Vertrauen untergraben wird. Diese Entwicklung droht sich durch das in der Forschung als „**Lügendividende**“ (liar's dividend) bezeichnete Phänomen weiter zuzuspitzen. Dieses ermöglicht es politischen Akteuren, die Echtheit unbequemer oder ihnen missliebiger Fakten, unter dem Vorwand, dass es sich um ein mögliches Deepfake handeln könnte, in Zweifel zu ziehen oder zu bestreiten. Der Einsatz von Deepfakes wird somit zu einem Instrument strategischer Leugnung und trägt zur Erosion des Vertrauens in Medien, Institutionen und letztlich auch in die Idee einer objektiven Wahrheit selbst bei.

In **Luxemburg** unterzeichneten die Parteien DP, LSAP, déi gréng, ADR, Déi Lénk, Piraten, Fokus und Volt im Januar 2023 ein gemeinsames **Wahlabkommen**, mit dem sie sich im Vorfeld der Europawahlen 2024 dazu verpflichteten, einen fairen und faktenbasierten Wahlkampf zu führen. Darin bekräftigten sie ausdrücklich ihre Absicht, soziale Netzwerke verantwortungsvoll nutzen und auf persönliche Angriffe, die Verbreitung von Falschinformationen, Diffamierungskampagnen sowie die Beschädigung oder Manipulation des Wahlkampfmaterials anderer Parteien verzichten zu wollen. Zwar enthält das Abkommen keine ausdrückliche Regelung zum Einsatz von Deepfakes, es hält jedoch fest, dass weder Bots noch automatisierte Propaganda- oder Manipulationskampagnen eingesetzt werden sollen. Vor diesem Hintergrund lässt sich ableiten, dass auch KI-gestützte Technologien wie Deepfakes nicht für Zwecke der Desinformation genutzt werden sollen.

Herstellung illegaler Inhalte

Dazu zählen insbesondere sogenannte „**Deepnudes**“, also **pornografische Inhalte**, die mithilfe künstlicher Intelligenz ohne das Einverständnis der betroffenen Personen erzeugt werden.

Die Herstellung, der Besitz oder die Verbreitung von Darstellungen sexuellen Missbrauchs von Minderjährigen ist strikt verboten und strafbar, ungeachtet dessen, ob diese mithilfe künstlicher Intelligenz erstellt wurden oder nicht.

Kombination mehrerer böswilliger Taktiken

Deepfakes werden nur selten einzeln verbreitet. In der Regel sind sie Teil **hybrider Kampagnen**, in denen verschiedene KI-gestützte Methoden miteinander kombiniert werden, um Reichweite und Wirkung gezielt zu maximieren.

Auch in **Luxemburg** sind Deepfakes von unter anderem Premierminister Luc Frieden, Bürgermeisterin der Stadt Luxemburg, Lydie Polfer, dem scheidenden Großherzog Henri sowie des ehemaligen Premierministers Jean-Claude Juncker in den sozialen Netzwerken zirkuliert. In den manipulierten oder vollständig KI-generierten Videos und Artikeln sprachen diese scheinbar

Anlageempfehlungen aus und forderten dazu auf, über bestimmte Kryptoplattformen zu investieren. Um die Glaubwürdigkeit dieser Betrugsmaschen zu erhöhen, wurden zudem Deepfakes der RTL- Journalistinnen Mariette Zenners, Caroline Mart und Lynn Cruchten sowie der Schauspielerin Désirée Nosbusch eingesetzt.

Die Betrüger kombinierten dabei mehrere Methoden, indem sie:

- seriöse und bekannte Nachrichtenseiten wie RTL Lëtzebuerg oder die „tagesschau“, samt Logos, Farbgestaltung und Seitenlayouts nachahmten;
- gefälschte Artikel mit reißerischen Überschriften erstellten, die angebliche sensationelle Enthüllungen versprachen;
- diese Inhalte über Fake-Profile sowie die kostenpflichtigen Werbedienste von Facebook und Instagram gezielt verbreiteten.

Was sieht das Gesetz vor?

Das Gesetz sieht Maßnahmen auf mehreren Ebenen vor:

- für **Entwickler** generativer KI-Modelle;
- für **Plattformen**, auf denen solche Inhalte verbreitet werden;
- für **Nutzerinnen und Nutzer (Anwendende)** dieser Werkzeuge, sofern deren Einsatz nicht ausschließlich auf den rein privaten Gebrauch beschränkt ist.

Pflichten der KI-Entwickler

Entwickler von KI-Systemen sind verpflichtet:

- sicherzustellen, dass jede Person, die mit einem KI-System interagiert, erkennen kann, dass es sich um eine Maschine und nicht um einen Menschen handelt;
- dafür zu sorgen, dass alle von der KI-erzeugten oder veränderten Inhalte als solche identifizierbar sind, etwa durch eine maschinenlesbare Kennzeichnung. Dazu zählen beispielsweise unsichtbare Wasserzeichen oder Metadaten wie Erstellungs- oder Änderungsdatum, das verwendete Modell oder ähnliche Informationen.

Um die Erstellung schädlicher Deepfakes zu verhindern, sind insbesondere folgende Vorkehrungen zu treffen:

- die Einrichtung von Eingabe- und Ausgabefiltern, um die Erstellung schädlicher Inhalte zu verhindern, insbesondere solcher, die Minderjährige betreffen oder sogenannte Deepnudes darstellen;
- die Durchführung regelmäßiger Risikobewertungen generativer KI-Systeme sowie die Prüfung der Rechtmäßigkeit der erzeugten Inhalte.

Pflichten der Plattformen

Plattformen sind verpflichtet:

- im Rahmen ihrer Moderationspflicht rechtswidriger Inhalte klare Regeln für die Erstellung und Verbreitung synthetischer Inhalte festzulegen;

- wirksame Sanktionen gegen Nutzerinnen und Nutzer vorzusehen, die gegen diese Regeln verstoßen, etwa durch die Sperrung oder Löschung von Konten;
- Möglichkeiten zu schaffen, potenziell rechtswidrige Deepfakes zu melden.

Plattformen sind zudem dazu angehalten:

- Werkzeuge bereitzustellen, mit denen kenntlich gemacht werden kann, ob ein Inhalt verändert oder mithilfe von KI erzeugt wurde, etwa durch einen entsprechenden Hinweis unter den betreffenden Beiträgen.

Pflichten der Nutzerinnen und Nutzer

Nutzerinnen und Nutzer sind dazu angehalten:

- klar kenntlich zu machen, wenn sie Deepfakes erzeugen, und davon abzusehen, schädliche oder illegale Deepfakes zu erstellen;
- verdächtige oder irreführende Inhalte melden.

Wie erkennt man Deepfakes?

Mit zunehmender Leistungsfähigkeit erzeugen KI-Modelle immer seltener offensichtliche Fehler, so dass es keine zuverlässigen visuellen oder akustischen Hinweise mehr gibt, um Deepfakes eindeutig als solche zu erkennen. Häufig lassen sich Deepfakes nur durch Analysen mithilfe spezialisierter Software oder durch aufwendige manuelle Untersuchungen, wie eine Bild-für-Bild-Analyse („Frame-für-Frame“) oder die Wiedergabe in Zeitlupe erkennen, wobei auch diese Verfahren bislang noch keine hundertprozentige Sicherheit bieten.

Das Fehlen „typischer“ KI-Fehler verleitet viele Menschen dazu, KI-Inhalte als authentisch zu bewerten.

Die Erkennung von Deepfakes stützt sich daher häufig weniger auf einzelne visuelle oder akustische Merkmale als vielmehr auf den **Kontext**, den **Inhalt** und mögliche **logische Widersprüche**. Zwar kann KI menschliche Stimmen, Gesten und Bewegungen überzeugend nachahmen, doch hat sie oft Schwierigkeiten mit physikalischen Gesetzen oder kausalen Zusammenhängen. Unlogisches Verhalten in einem Video, wie etwa eine Person, die regungslos bleibt, wenn ein Feuer ausbricht, oder ein Auto, das entgegen der Fahrtrichtung auf einem Radweg fährt, kann die Illusion eines echten Ereignisses brechen und darauf hinweisen, dass es sich um ein Deepfake handelt.

Was können Sie tun?

- **Bleiben Sie informiert.** Lesen Sie Fachartikel und abonnieren Sie Newsletter oder Podcasts zum Thema Künstliche Intelligenz. Da sich die Technologie rasant weiterentwickelt, gilt: Je besser sie verstehen, wie KI funktioniert, desto leichter erkennen Sie manipulierte Inhalte.
- **Achten Sie auf KI-Kennzeichnungen.** Hinweise darauf, dass ein Inhalt mithilfe von KI erstellt wurde, können je nach Plattform unterschiedlich gestaltet und platziert sein (z.

B. Farbe, Symbol, Text oder Position). Häufig finden sich entsprechende Angaben in der Bild- oder Videobeschreibung. Achten Sie auf Begriffe wie **#AI**, **#GeneratedWithAI**, **#deepfake** oder auf das Logo des verwendeten Tools (z. B. *Sora* von OpenAI).

- **Teilen Sie Inhalte nicht unbedacht.** Jeder Like, Kommentar oder jedes Teilen trägt zur Verbreitung potenziell manipulierter Inhalte bei. Wenn Sie Zweifel an der Echtheit eines Inhalts haben, verzichten Sie darauf, diesen zu teilen.
- **Prüfen Sie die Medienberichterstattung.** Seriöse Medien berichten in der Regel schnell über relevante Ereignisse. Findet sich eine Meldung ausschließlich auf einzelnen Social-Media-Kanälen und sind sonst nirgends auffindbar, handelt es sich wahrscheinlich um einen Fake. Prüfen Sie Informationen möglichst immer anhand mehrerer Quellen. Nutzen Sie dafür auch Faktencheck-Angebote wie **EDMO Belux** oder **RTL Faktencheck**.
- **Bewahren Sie eine kritische Haltung.** Schätzungen zufolge könnten bis 2026 rund 90 % der Online-Inhalte KI-generiert sein. Bleiben Sie daher aufmerksam, ohne jedoch in eine pauschale Skepsis gegenüber verlässlichen journalistischen oder offiziellen Quellen zu verfallen. In einer zunehmend unübersichtlichen Informationslandschaft ist Vertrauen in transparente, verantwortungsvolle und glaubwürdige Kommunikation entscheidend.

Nützliche Links und Ressourcen

- Auf letzfin.lu, der Referenzplattform für Finanzbildung der luxemburgischen Finanzaufsichtsbehörde Commission de Surveillance du Secteur Financier (CSSF), finden Sie wichtige Informationen und praktische Hinweise zum Schutz vor Finanzbetrug. Der Bereich **„**Vorsichtsmaßnahmen**“** sensibilisiert die Öffentlichkeit für die häufigsten Betrugsmaschen, darunter KI-gestütztes Phishing, Identitätsdiebstahl und gefälschte Websites. Folgen Sie auch **@letzfin** auf Instagram, um informiert zu bleiben.
- Das Nationale Kompetenzzentrum für Cybersicherheit (NC3), angesiedelt beim Luxembourg House of Cybersecurity (LHC), bietet Unternehmen und Institutionen Beratung sowie praktische Werkzeuge zur Stärkung ihrer digitalen Sicherheit, auch im Hinblick auf Angriffe, bei denen künstliche Intelligenz eingesetzt wird.
- Im Falle von Online-Betrug unterstützt die Website cyberfraud.lu dabei, die richtigen Schritte einzuleiten und die zuständigen Ansprechpartner zu finden. Diese Initiative wurde im Juni 2025 vom Luxembourg House of Cybersecurity (LHC) und der Association des banques et banquiers Luxembourg (ABBL) im Rahmen einer nationalen Sensibilisierungskampagne gestartet und wird von mehr als fünfzehn Partnerorganisationen unterstützt.
- Bei allgemeinen Fragen zur Cybersicherheit und Online-Sicherheit wenden Sie sich an die **BEE SECURE Helpline** unter **+352 8002 1234** oder über das Online-Formular.