Deepfakes – eine echte Bedrohung

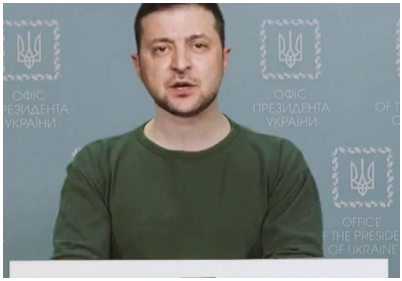
Verstehen, reagieren, sich schützen

Ihre Polizei und die Schweizerische Kriminalprävention (SKP) – eine interkantonale Fachstelle der Konferenz der kantonalen Justiz- und Polizeidirektorinnen und -direktoren (KKJPD), in Zusammenarbeit mit dem Institut de Lutte contre la criminalité économique (ILCE) der Haute école de gestion Arc, Neuenburg.

Deepfakes – ein boomendes Phänomen

Deepfakes (englisches Kofferwort aus *deep learning* und *fake)* sind Bild-, Video und Audio-Inhalte, die mit Techniken der künstlichen Intelligenz erzeugt oder abgeändert wurden. So werden meist Stimmen und äussere Eigenschaften von Menschen sehr realistisch imitiert.

Beispiele:



2022: Deepfake-Video, in dem der ukrainische Präsident Wolodimir Selenski seine Soldaten zur Kapitulation aufruft.

2023: Deepfake-Bild von Papst Franziskus mit weisser Daunenjacke.
(Bild: Mit der KI-Software Midjourney generierter Deepfake)

Deepfakes werden für Irreführung, Manipulation und Betrug verwendet. Sie lassen sich kaum von der Realität unterscheiden und verbreiten sich teils rasant über das Internet und die Sozialen Medien.

Konkret können Deepfakes folgenden Zwecken dienen:

- Verbreitung von Fehlinformationen
- Manipulation der öffentlichen Meinung
- Identitätsdiebstahl
- Betrug
- Schaffen und Verbreiten gefälschter pornografischer Bilder
- Rufschädigung
- Nachstellung, Mobbing

Fast unbegrenzte Möglichkeiten

Es gibt immer mehr Tools zur Schaffung von Deepfake-Inhalten, und sie sind immer einfacher zu haben. Damit lassen sich relativ einfach Bilder, Videos und Audioaufnahmen erstellen, welche die Realität verfälschen und manipulieren. Diese Inhalte werden über öffentliche und private Kanäle verbreitet (z. B. WhatsApp) und können in diversen Formen unterschiedlichen Zwecken dienen, z. B.:

- eine öffentliche Persönlichkeit in einen erfundenen Zusammenhang stellen oder eine gefälschte Rede halten lassen,
- eine prominente Person ohne Einwilligung in einer Werbung verwenden oder in Szene setzen,
- ein geschichtliches Ereignis oder eine tagesaktuelle Situation abändern,
- für Online-Profile fiktive Gesichter und Personen erstellen,
- Details an Bildern oder Videos verfälschen,
- Menschen in absurden oder lächerlichen Situationen darstellen,
- aus realen Bildern und Videos fiktive Szenen erstellen,
- zu Betrugszwecken die Stimme einer Person nachahmen,
- Bilder von Bekannten benutzen, um pornografische Inhalte zu erstellen,
- eine Stimmerkennung oder Videoidentifikation überlisten,
- eine Desinformationskampagne fördern
- und vieles andere mehr.

Persönlichkeitsverletzungen und Desinformation

Zur politischen Einflussnahme wird KI auf verschiedene Weise eingesetzt. Manchmal geht es darum, mit gezielter Falschinformation oder mit bewusst einseitiger Kommunikation in den sozialen Netzwerken Meinungen und somit auch Abstimmungen und Wahlen zu beeinflussen. Dazu werden Texte, aber auch Bilder und Videos gefälscht. In anderen Fällen steht die Provokation im Mittelpunkt. Auch in der Schweiz gab es schon Fälle von gefälschten Videos mit unwahren Aussagen zum Zweck der Verunglimpfung.

Beispiel 1

Online-Anlagebetrug

Auf einer Onlineplattform entdeckt Paul den Hinweis auf einen «Blick»-Artikel. Roger Federer plaudert aus dem Nähkästchen und verrät, wie er sein Vermögen vergrössern konnte. Paul wird demnächst pensioniert, aber seine Pension ist sehr knapp. Das wäre eine ideale Gelegenheit, seine Finanzlage zu verbessern. Daher klickt er auf den Link, kommt auf die «Blick»-Seite und sieht sich das Interview an. Zu seiner Freude gibt Roger Federer sogar den Link auf die Investitionsplattform bekannt. Mit nur CHF 250.– kann er dort bereits einsteigen. Alles gut?

Nein! Es ist alles falsch. Das Inserat wurde von Kriminellen geschaltet, die vermeintliche «Blick»-Seite wurde nachgebildet und hat mit der Zeitung «Blick» nichts zu tun und Roger Federers Finanztipps sind komplett erfunden. Sein Bild und auch sein Video wurden mit KI gefälscht. Weshalb sollte Roger Federer auch öffentlich über seine Investitionen sprechen? Es gibt übrigens Kriminelle, die diese kritischen Fragen schon vorweggenommen haben. Daher gibt es nun auch Inserate und vermeintliche Artikel zu Prominenten, die zusammengeschlagen wurden, weil sie ihre Investitionsgeheimnis ausplauderten. Und Sie haben es erkannt: Auch das ist ein Fake!

Beispiel 2

Romance Scam

Anita hat nach längerer Suche im Internet endlich ihren Traumpartner Rolf gefunden. Er ist Ingenieur und zurzeit auf einer Bohrinsel im Atlantik tätig. Rolf hat sich unsterblich in Anita verliebt, darum würde er sie gerne so rasch als möglich besuchen, aber die Umstände lassen es nicht zu. Sie chatten täglich und schicken sich regelmässig Bilder. Da wird Rolf krank. Zur Behandlung muss er zurück aufs Festland. Aber seine Konten sind gesperrt, und daher bittet er Anita um Hilfe. Sobald er genesen ist, wird er sie auch endlich besuchen können. Alles gut?

Nein! Es ist alles falsch. Rolf ist in Wirklichkeit ein Krimineller, der ganz anders aussieht, eine andere Sprache spricht, an einem anderen Ort wohnt und nur am Geld interessiert ist. Mit Hilfe von KI-Assistenten kann er seine Chats in die gewünschte Sprache übersetzen. Das Bild hat er aus dem Internet und passte es mit KI den gewünschten Situationen an. Und Sie ahnen es schon: Er wird Anita nie besuchen.

Auf Deepfakes reagieren

Deepfakes und andere manipulierte Inhalte können sogar für Spezialistinnen und Spezialisten äusserst realistisch erscheinen. Darum sind Wachsamkeit und gute Reflexe notwendig, bevor eine Information geglaubt und geteilt wird. Bei der Verwendung von Tools, die Deepfakes angeblich sicher entlarven sollen, ist Vorsicht geboten. Um aus den täglich konsumierten Informationen die manipulierten Inhalte herauszufiltern, muss darum richtiges Verhalten eingeübt werden.

Denken – Prüfen – Melden

1. Denken: Kritisch sein

- Misstrauen Sie Inhalten, die auf starke Gefühle zielen.
- Reagieren Sie nicht übereilt darauf.
- Fragen Sie sich: Wer hat diesen Inhalt publiziert? Wieso?

Oft wollen Deepfakes Angst, Wut und Überraschung wecken, um irrationale Reaktionen auszulösen.

2. Prüfen: Faktencheck vornehmen

- Wer ist die Autorenschaft des Inhalts?
- Wie zuverlässig ist die Quelle?
- Erscheint die Information in mehreren anerkannten Medien?

Echte Informationen werden im Allgemeinen von mehreren zuverlässigen Quellen verbreitet.

3. Melden: Deepfakes anprangern

- Teilen Sie Deepfakes nicht mit anderen.
- Melden Sie sie bei der Plattform, auf der sie erscheinen.
- Informieren Sie Ihre Umgebung, damit sie nicht weiter verbreitet werden.

Wenn Sie Deepfakes melden und nicht verbreiten, tragen Sie zur Bekämpfung des Phänomens bei.

Und wenn Sie Opfer werden?

Deepfakes können ganz gezielt gegen Einzelpersonen gerichtet sein. Mit Eingriffen in die Privatsphäre und wirtschaftlichen Schäden können die Folgen schwerwiegend sein – speziell bei Identitätsdiebstahl, Nachstellung (Mobbing), Verbreitung von pornografischem Material und Betrug. Darum ist es sehr wichtig, Reichweite und Wirkung solcher Inhalte einzuschränken.

- **Verbreiten Sie keine Deepfakes,** auch nicht zur Warnung Ihres Umfelds.
- **Behalten Sie alle Beweismittel:**
  - Screenshots
  - Links zu den Inhalten
  - Korrespondenz
- **Melden Sie Deepfakes** auf der betroffenen Plattform und verlangen Sie ihre Löschung.

Verschiedene Straftatbestände sind möglich, namentlich:

- Betrug (Art. 146 StGB)
- Üble Nachrede (Art. 173 StGB) und Verleumdung (Art. 174 StGB)
- Identitätsmissbrauch (Art. 179^{decies} StGB)
- Nachstellung (Art. 181b StGB)
- Pornografie (Art. 197 StGB), Unbefugtes Weiterleiten von nicht öffentlichen sexuellen Inhalten (Art. 197a StGB) und Sexuelle Belästigungen (Art. 198 StGB)

Aufgrund des Persönlichkeitsschutzes sind auch zivilrechtliche Schritte möglich (Art. 28, 28a und 28b ZGB).

Zögern Sie nicht, bei einer Rechtsverletzung Anzeige bei Ihrer Kantonspolizei zu erstatten.

Weitere Informationen und Beratung

Falls Sie Zweifel oder einen Verdacht auf Deepfakes haben, informieren Sie sich auf der Website des Bundesamts für Cybersicherheit (www.ncsc.admin.ch) oder auf cybercrimepolice.ch. Hier finden Sie aktuelle Informationen, Präventionsratschläge und Praxistipps sowie Angaben, wie Sie Anzeige erstatten können.

Weitere Informationen und Materialien zur Cybersicherheit erhalten Sie auch auf der Website der Schweizerischen Kriminalprävention (www.skppsc.ch).

Mit Deepfakes konfrontiert?

→ Kritisch **denken**
→ Fakten **prüfen**
→ Deepfakes **melden**

Jeder Faktencheck trägt zu weniger Fehlinformation bei. Verbreiten Sie keine Deepfakes!

Folgen Sie uns auf:

Facebook: @Schweizerische Kriminalprävention **Instagram:** @skppsc_schweiz
LinkedIn: @Schweizerische Kriminalprävention **YouTube:** @SKPPSCSCP

Für weitere Informationen: www.skppsc.ch

SKPPSC

Schweizerische Kriminalprävention
Haus der Kantone
Speichergasse 6
3001 Bern

www.skppsc.ch

Juni 2026