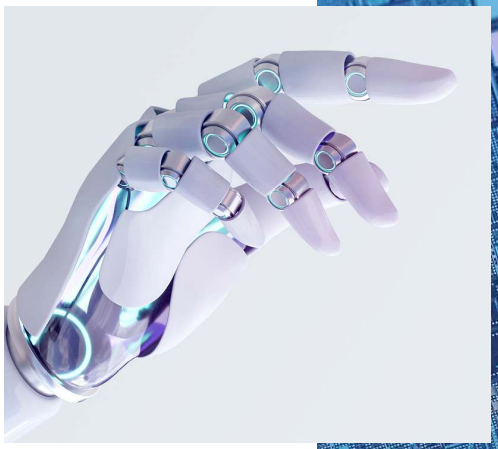
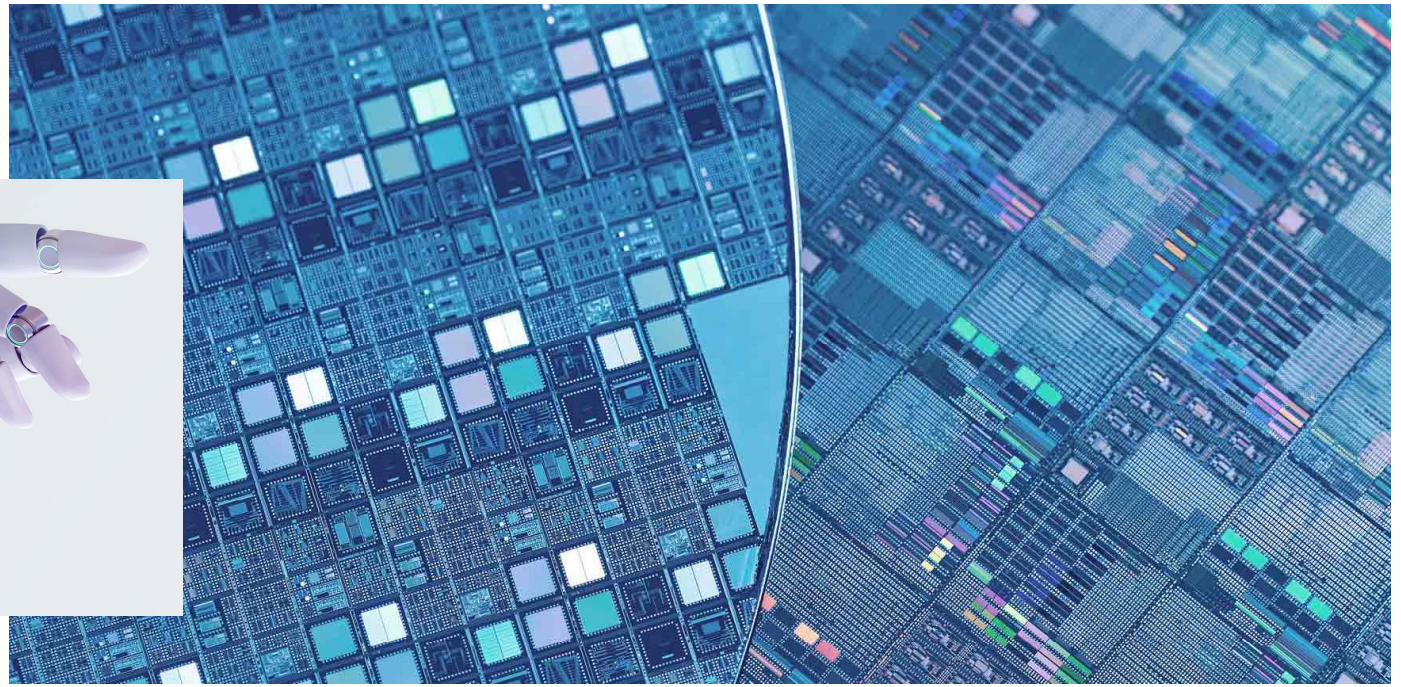
Vorausschauend
seit Generationen

Deepfakes – Die neuen Herausforderungen der digitalen Welt





digital-liechtenstein.li
Cybersecurity-Konferenz 2024

Deepfakes – Die neuen Herausforderungen der digitalen Welt

Einleitung

Sicherheitsvorfälle mit Einbezug von Deepfakes

Schutzmassnahmen

Deepfakes – Die neuen Herausforderungen der digitalen Welt

Einleitung

Sicherheitsvorfälle mit Einbezug von Deepfakes

Schutzmassnahmen

Übersicht Deepfake

- Was sind «Deepfakes»?
  - «Deepfakes sind realistisch wirkende Medieninhalte, die durch Techniken der künstlichen Intelligenz (AI) abgeändert, erzeugt oder verfälscht werden.» (Wikipedia)

- AI-Technologien, die Bilder, Texte, Audio und Videos erzeugen können, eröffnen neue kreative und kommerzielle Möglichkeiten. Gleichzeitig steigt das Risiko ihres Missbrauchs, etwa für Manipulation, Betrug oder Belästigung.

- Identitätsdiebstahl und Betrug existieren schon bevor es AI gab. Durch den breiteren Zugang dieser Technologien lassen sich gefälschte Inhalte täuschend echt darstellen.

- Missbrauch von AI
  - **Ausnutzung der AI-Fähigkeiten**: Erstellen realistischer Medieninhalte zur Täuschung.
  - **Kompromittierung der AI-Systeme**: Entfernen von Sicherheitsmassnahmen oder Einsatz gezielt manipulierter Eingaben, um Fehlfunktionen herbeizuführen.

Disclaimer:
Die Bilder und Audiodateien in dieser Präsentation sind KI-generierte Deepfakes, die ausschliesslich zu Illustrations- und Bildungszwecken erstellt wurden. Sie sind nicht echt und sollten nicht als echte Darstellungen von Personen oder Ereignissen interpretiert werden.
Bitte betrachten Sie den Inhalt kritisch und seien Sie sich des Manipulationspotenzials in digitalen Medien bewusst. Vielen Dank für Ihr Verständnis.

Deepfake Beispiel: Donald Trump

https://www.bbc.com/news/world-us-canada-68440150

Deepfakes – Die neuen Herausforderungen der digitalen Welt

Einleitung

Sicherheitsvorfälle mit Einbezug von Deepfakes

Schutzmassnahmen

Vorfall Schweiz: CEO-Betrug 2.0

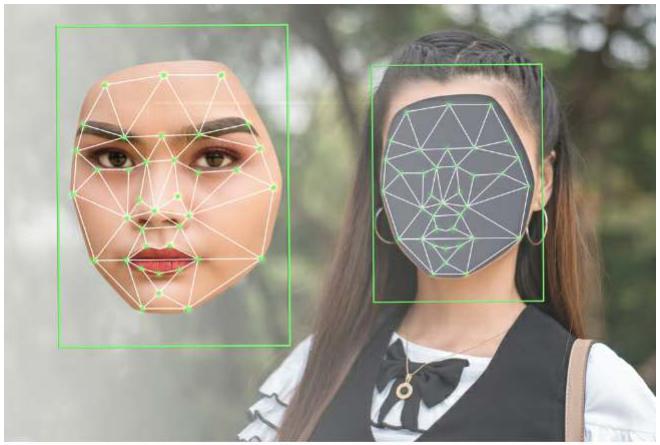
- Der klassische CEO-Betrug
  - **Szenario**: Eine dringende Zahlungsaufforderung vom angeblichen Chef trifft bei der Finanzabteilung ein. Der Chef betont die Dringlichkeit und droht mit ernsthaften Konsequenzen, falls die Zahlung nicht erfolgt. Rückfragen bleiben unbeantwortet.
  - **Vorgehen**: Betrüger durchsuchen Unternehmenswebsites und LinkedIn-Profile nach Informationen über CEOs und Finanzverantwortliche. Sie senden dann gefälschte E-Mails mit dringenden Zahlungsanweisungen.
  - **Erkennungsmerkmale**: Diese Angriffe sind meist nicht sehr ausgereift und leicht zu durchschauen. Texte sind unspezifisch und oft identisch.

- Neue Bedrohung durch Deepfake:
  - **Beispiel eines tatsächlichen Vorfalls, welcher dem Bundesamt für Cybersicherheit (BACS) gemeldet wurde**:
    Ein aktueller Fall zeigt eine raffiniertere Methode. Ein Finanzverantwortlicher wurde von einem angeblichen Anwalt zu einer Videokonferenz mit dem Chef eingeladen. In der Konferenz sah und sprach er mit einem Deep-Fake-Video des Chefs, das mittels Künstlicher Intelligenz erstellt wurde. Der vermeintliche Chef versuchte, die Mobiltelefonnummer des Finanzverantwortlichen zu erhalten und Finanztransaktionen auszulösen.
  - **Erkennung:** Der Betrug flog schnell auf, da die Kleidung und Stimme des Chefs nicht authentisch wirkten.
  - Dieser Vorfall zeigt, dass Betrüger zunehmend Künstliche Intelligenz nutzen, auch wenn die Umsetzung noch nicht perfekt ist
  - https://www.ncsc.admin.ch/ncsc/de/home/aktuell/im-fokus/2024/wochenrueckblick_14.html

Vorfall International: ‘Everyone looked real’

- Ein multinationales Unternehmen verlor HK$200 Millionen (US$25,6 Millionen) durch einen Betrug mit Deepfake-Technologie.
- Die Betrüger verwendeten Deepfake-Technologie, um überzeugende Darstellungen des CFO und anderer Mitarbeiter des Unternehmens zu erstellen.
- Das Opfer wurde während einer Videokonferenz mit mehreren Teilnehmern getäuscht, bei der alle außer dem Opfer Deepfake-Darstellungen waren.
- Die Betrüger nutzten öffentlich zugängliche Video- und Audioaufnahmen, um die Deepfakes zu erstellen.
- Das Opfer erhielt eine Phishing-Nachricht von dem vermeintlichen britischen CFO, in der eine geheime Transaktion gefordert wurde.
- Das Opfer glaubte, dass der Deepfake-CFO und die anderen Teilnehmer echt seien, und folgte den Anweisungen, HK$200 Millionen auf fünf Hongkonger Bankkonten zu überweisen.
- Der Betrug dauerte etwa eine Woche, bevor das Opfer den Betrug erkannte, als es eine Anfrage an die Unternehmenszentrale stellte.

- https://www.scmp.com/news/hong-kong/law-and-crime/article/3250851/everyone-looked-real-multinational-firms-hong-kong-office-loses-hk200-million-after-scammers-stage



Deepfake technology allows for face swapping and matching of facial movements with a different person. Photo: Shutterstock

Vorfall International – Fake Software Engineer aus Nordkorea

- **Beschreibung des Vorfalls**
  - Die amerikanische Cybersicherheitsfirma „KnowBe4“ stellte einen Software Engineer an, welcher sich als nordkoreanischer Hacker entpuppte.
  - Dieser nutze die gestohlene Identität einer US-Person und erstellte ein AI-generiertes Profilbild, welches er für die Jobinterviews verwendete.
  - Aufgeflogen war er weil er direkt nach seiner Anstellung versuchte eine Malware auf dem Unternehmensgerät zu installieren.

- **Bewerbungsprozess**: Backgroundchecks, Referenzverifizierungen und vier Video-Interviews wurden durchgeführt, um die Identität zu bestätigen.

- https://blog.knowbe4.com/how-a-north-korean-fake-it-worker-tried-to-infiltrate-us

The original stock picture (left) and an AI fake (right) used by a North Korean threat actor who posed as a U.S.-based software engineer and was hired by the cyber firm KnowBe4. (Photo credit: KnowBe4)

Deepfakes – Die neuen Herausforderungen der digitalen Welt

Einleitung

Sicherheitsvorfälle mit Einbezug von Deepfakes

Schutzmassnahmen

Schutzmassnahmen

- Bleiben Sie kritisch
  - Handelt es sich um einen offiziellen Kommunikationskanal?
  - Ist es plausibel?

- Mitarbeitende sensibilisieren
  - Mitarbeitende bezüglich CEO-Fraud sensibilisieren. Insbesondere Mitarbeitende in Finanzabteilungen oder in Schlüsselpositionen
  - Sensibilisieren Sie Mitarbeitende dahingehend, dass gezielte Angriffe mit öffentlich verfügbaren Informationen(z. B. aus LinkedIn) durchgeführt werden können.

- Vorsicht bei ungewöhnlichen Zahlungsaufforderungen

- Vorsicht bei Preisgabe von internen Informationen solange die Identität des Gegenübers nicht klar ist (weder per E-Mail noch per Telefon)

- Rückfragen über einen gesicherten 2. Kanal

- Einhaltung der Prozesse, welche den Zahlungsverkehr betreffen (z. B. Vier-Augen-Prinzip, Kollektivunterschrift zu zweien).