Der Zugang zur Deepfake-Software ist sehr einfach, manchmal sogar kostenlos. Foto: Imago

Wie Firmen sich gegen Deepfakes wehren können

Künstliche Intelligenz Die Bedrohung durch Betrug mit KI nimmt zu und führt in der Wirtschaft zu Milliardenverlusten. Unternehmen müssen sich organisatorisch wappnen.

Bernhard Kislig

Anfang 2024 erhielt eine Buchhalterin in der Hongkonger Niederlassung des international tätigen Planungs- und Beratungsunternehmens Arup Group per Videoanruf Anweisungen ihres Finanzleiters und weiterer Führungskräfte aus dem Hauptsitz in London. Sie forderten sie auf, eine Überweisung an ein externes Konto zu tätigen. Was sie nicht wusste: Alle Personen waren mit KI generierte Fälschungen. Betrüger hatten deren Erscheinung, Mimik und Stimme täuschend echt simuliert. Die Mitarbeiterin überwies umgehend 25 Millionen US-Dollar, die sofort weitertransferiert und somit verloren waren.

Dieses Beispiel für Deepfake-Betrug per Live-Video zeigt das Schadenspotenzial dieser Technologie. Deepfakes sind mit künstlicher Intelligenz erzeugte Medieninhalte – meist geht es um Videos, Audiodateien oder Bilder. Das «Fake» steht für Fälschung.

Neue Dimension des Betrugs

Solche Vorfälle unterstreichen die neue Dimension des Betrugs: Die Opfer handeln selbst, überzeugt von der Echtheit der Manipulation. So hebeln Betrüger die klassischen Sicherheitsvorkehrungen aus. Unternehmen stellt das vor neue Herausforderungen.

Das Wirtschaftsprüfungs- und Beratungsunternehmen Deloitte schätzt, dass allein in den USA die Schäden durch Deepfake-Betrug bei Unternehmen von 12,3 im Jahr 2023 auf 40 Milliarden Dollar im Jahr 2027 zunehmen. Das in der Identitätsverifikation tätige Unternehmen Regula führte in den Jahren 2022 und 2024 internationale Umfragen zu solchen Betrugsfällen durch. Das Ergebnis: Der durchschnittliche Schaden durch solche Betrugsfälle je Unternehmen stieg in diesem Zeitraum von 230’000 auf 450’000 Dollar. Bei Finanzinstituten lag er höher.

Der Grund für den Anstieg ist der einfache Zugang zur Technologie. Die dafür notwendige Software ist in der Anwendung nicht kompliziert – viele Produkte stehen sogar kostenlos zur Verfügung. Und dank zunehmend ausgefeilter Technologie wirken die Ergebnisse je länger je überzeugender. So ist es heute beispielsweise möglich, nicht nur Stimme und Tonlage, sondern beispielsweise auch den Akzent einer fremden Sprache nachzuahmen. Bild und Ton schaffen Vertrauen. Zudem provozieren Betrüger in der Regel Zeitdruck und emotionalen Stress: Die angeblichen Vorgesetzten beschreiben eine Notsituation, die sofortiges Handeln erfordert. Angestellte zögern dann eher, die dringenden Anweisungen zu hinterfragen.

Die schnelle Entwicklung und Zugänglichkeit von Deepfake-Technologien stellt traditionelle Sicherheits- und Kontrollsysteme vor Herausforderungen. Bislang verliessen sich Unternehmen auf etablierte Verfahren zur Verifizierung von Identitäten und zur Authentifizierung von Kommunikationen. Dazu gehören biometrische Merkmale, visuelle Dokumentenprüfungen oder Stimmabdrucksysteme beim Telefonbanking. Deepfakes zielen jedoch genau auf diese Schwachstellen ab. Künstlich erzeugte Gesichter können Passfotoqualität erreichen und Blinzeln oder Kopfbewegungen simulieren, wodurch einfache Checks umgangen werden.

Künstlich erzeugte Gesichter können Passfotoqualität erreichen und Blinzeln oder Kopfbewegungen simulieren.

Rechtsanwalt Fabian Teichmann von der Teichmann International AG hat sich in einem Beitrag in der juristischen Fachzeitschrift «Jusletter» mit der Frage auseinandergesetzt, wie sich Unternehmen gegen Betrugsversuche mit Deepfakes wehren können. Er nennt unter anderem die nachfolgenden organisatorischen Massnahmen und technischen Hilfsmittel:

— **Striktes Vieraugenprinzip**: Bei ungewöhnlichen Zahlungsaufforderungen sollte es Pflicht sein, bei einer anderen Person eine Zusage oder über andere Kanäle eine schriftliche Bestätigung einzuholen. Ein Rückruf darf nur über eine bekannte Nummer erfolgen und nicht über einen Kontakt, den Vorgesetzte im Videocall nennen. Unternehmen können zudem für Notfälle auch ein Codewort vereinbaren, das Betrüger nicht kennen.

— **Mitarbeiterschulungen**: Insbesondere Mitarbeitende, die Zahlungen auslösen können, sollten regelmässig Schulungen zu Deepfake-Betrugsmethoden besuchen. Simulationsübungen können das Bewusstsein schärfen. Mitarbeitende müssen sich zutrauen, auch Anweisungen von ihren Vorgesetzten zu verifizieren, und die dafür notwendige Rückendeckung erhalten.

— **Zurückhaltung bei sozialen Medien**: Schon wenige gesprochene Worte können ausreichen, um mit KI eine Stimme zu imitieren. Es ist ratsam, wenn Verantwortungsträger in sozialen Medien mit Video- und Bildmaterial zurückhaltend umgehen.

— **Erkennungssoftware**: Software, die Videos und Bilder auf typische Artefakte oder Inkonsistenzen scannt, kann helfen, Verdachtsmomente auszulösen. Diese Software kann Deepfakes aber nicht mit absoluter Sicherheit erkennen.

Prüfmethoden ergänzen

Neben dem «CEO-Fraud», bei dem ein angeblicher Geschäftsführer per Videocall eine Zahlung veranlasst, besteht ein weiteres Risiko für Finanzdienstleister. Bei der Online-Kontoeröffnung können Kriminelle mit KI täuschend echt gefälschte Ausweisdokumente erstellen. Auch Videoaufnahmen beim Identifizierungsprozess lassen sich künstlich erstellen.

Banken müssen traditionelle Prüfmethoden ergänzen, um durch KI erstellte Fälschungen zu erkennen. Teichmann empfiehlt auch technische Hilfsmittel, die künstlich erstellte Ausweisfotos erkennen können. Diese erkennen etwa ungewöhnliche Artefakte oder stellen fest, wenn ein Bild in einer öffentlichen Datenbank vorhanden ist. Letzteres kann ein Indiz für eine Fälschung sein. Hilfreich ist auch eine erweiterte Dokumentenprüfung, die etwa Hologramme auf Ausweispapieren einbezieht. Dies kombiniert mit einer Zweitprüfung durch Menschen bei Unstimmigkeiten. Mehr Sicherheit schafft im Zweifelsfall eine zusätzliche Identifikation am Schalter.