Voice-Cloning und KI-Avatare: Digitale Betrugsmöglichkeiten werden professioneller

12.8.2026 - | Gesamtverband der Deutschen Versicherungswirtschaft

Betrugsanrufe und Meetings mit gefälschten Identitäten werden immer professioneller. Der GDV gibt einen Überblick über die Betrugsmaschen und zeigt mögliche Sicherheitsvorkehrungen auf.

Ein Anruf vom Chef, der um eine dringende Überweisung bittet oder ein Video-Call mit dem Vorstand, der eine Ausnahme vom üblichen Freigabeprozess fordert. Was früher eine plumpe Betrugsmasche war, klingt heute täuschend echt. Voice-Cloning und KI-Avatare haben synthetische Stimmen und Gesichter in den Alltag von Betrügern integriert. Oft genügt ein dringender Anruf zur falschen Zeit, um etablierte Kontrollen auszuhebeln. Der Gesamtverband der Deutschen Versicherungswirtschaft (GDV) klärt über die verbesserten Betrugsmaßnahmen auf und gibt einen Überblick über mögliche Sicherheitsvorkehrungen.

Was heute technisch möglich ist

Aus wenigen Minuten Sprachaufnahme lässt sich ein Modell einer Stimme erzeugen. Manche Anbieter kommen inzwischen mit fünf bis 15 Sekunden an Audiomaterial aus. Sie unterstützen über 80 Sprachen und liefern die ersten Audioausgaben innerhalb von rund 90 Millisekunden. Das Modell kann anschließend beliebigen Text in der geklonten Stimme sprechen. Es kann auch eine vorhandene Aufnahme in Klang und Charakteristik einer anderen Person umwandeln.

Auch Echtzeit-Verfahren sind verfügbar. Die Stimme lässt sich während eines Telefonats oder Videogesprächs mit geringer Verzögerung verändern. KI-Avatare gehen noch weiter. Sie synchronisieren Lippenbewegungen, erzeugen Gestik und reagieren live auf Rückfragen. Ein Angreifer spielt damit nicht nur eine vorbereitete Nachricht ab. Er führt ein echtes Gespräch.

Als Ausgangsmaterial genügen oft öffentlich zugängliche Aufnahmen. Interviews, Podcasts, Konferenzvideos oder Beiträge in sozialen Netzwerken liefern ausreichend Stimmproben für ein überzeugendes Klonmodell.

Vom Anruf zum gefälschten Meeting

Das klassische Betrugsmuster bleibt erkennbar, auch wenn sich die Mittel weiterentwickelt haben. Ein Angreifer recherchiert das Organigramm und den Kommunikationsstil einer Zielperson. Er sammelt Bild- und Audiomaterial und imitiert eine Stimme oder ein Gesicht. Dann erzeugt er Zeitdruck oder eine angebliche Ausnahmesituation. Die Rückruf- oder Freigabekontrolle wird als geschäftlich unmöglich dargestellt.

Wie weit das gehen kann, zeigte ein Fall beim britischen Architektur- und Ingenieurbüro Arup Anfang 2024. Ein Mitarbeiter der Hongkonger Niederlassung erhielt eine Nachricht, die angeblich vom Finanzchef der britischen Zentrale stammte. Sie kündigte eine dringende, vertrauliche Transaktion an. Trotz anfänglicher Zweifel nahm er an einer Videokonferenz teil. Dort saßen vermeintlich mehrere bekannte Kollegen, darunter der Finanzchef selbst. In Wirklichkeit waren alle Teilnehmer Deepfakes. Über eine Woche hinweg überwies der Mitarbeiter in 15 Transaktionen rund 23 Millionen Euro auf die Konten der Täter. Erst ein Rückruf bei der Unternehmenszentrale in

Großbritannien brachte den Betrug ans Licht.

Warum Erkennung allein nicht reicht

Es bestehen weiterhin Anhaltspunkte für Fälschungen. Unnatürliche Betonung, monotone Sprechweise, falsch ausgesprochene Namen oder unpassende Hintergrundgeräusche können auffallen. Bei Video-Avataren verraten manche Fälschungen ungenaue Lippensynchronisation oder eine eingeschränkte Mimik.

Diese Hinweise bleiben unsicher. Schlechte Kameraqualität oder eine wacklige Verbindung lassen auch echte Gespräche künstlich wirken. Die Qualität synthetischer Inhalte verbessert sich laufend. Wer sich allein auf das Erkennen von Fälschungen verlässt, baut auf schwindende Sicherheit. Belastbarer ist die Überprüfung der Handlung selbst als die Beurteilung der Bild- oder Tonqualität.

In der Praxis haben sich einige Maßnahmen bewährt:

- Rückruf über eine bereits bekannte Nummer statt über die Nummer aus der Nachricht
- Bestätigung über einen zweiten, unabhängigen Kommunikationsweg
- Vier-Augen-Prinzip bei Zahlungen und Änderungen von Bankdaten
- Keine Aufhebung bestehender Freigabeprozesse wegen Zeitdrucks
- Vereinbarte Codewörter für besonders kritische Vorgänge
- Mehrfaktor-Authentifizierung, die nicht allein auf Stimme oder Gesicht beruht

Langfristig gewinnen kryptografische Herkunftsnachweise an Bedeutung. Digitale Signaturen im Aufnahmeprozess zeigen später, wer eine Aufnahme erstellt hat. Sie zeigen auch, ob die Aufnahme nachträglich verändert wurde. Ein solcher Nachweis beweist nicht automatisch, dass ein Inhalt wahr ist. Er hilft aber dabei, Herkunft und Integrität einzuordnen.

Wie sich der Versicherungsschutz aufteilt

Bei Voice-Cloning-Betrug ist die Vertrauensschadenversicherung die wesentliche Deckung. Sie ist genau für solche Fälle gemacht. Ein Mitarbeitender wird durch eine gefälschte Stimme, ein gefälschtes Video oder eine betrügerische E-Mail dazu gebracht, selbst eine Überweisung auszulösen. Die Vertrauensschadenversicherung ersetzt in diesen Fällen das verlorene Geld.

Eine Cyberversicherung spielt dagegen nur in Ausnahmefällen eine Rolle. Verrät eine Mitarbeitende oder ein Mitarbeiter infolge eines gefälschten Anrufs seine Zugangsdaten, ist das für sich genommen noch kein Cyberschaden. Erst wenn Angreifer mit diesen Zugangsdaten in Systeme eindringen und dort selbst Schaden anrichten, etwa durch Sperrung, Diebstahl, Manipulation oder Verschlüsselung von Unternehmensdaten, greift die Cyberversicherung. Sie deckt vor allem die technischen Folgen ab, etwa die Datenwiederherstellung, den Neuaufbau von Systemen, die Kosten für Forensik sowie eine mögliche Betriebsunterbrechung.

https://www.gdv.de/gdv/themen/digitalisierung/voice-cloning-und-ki-avatare-digitale-betrugsmoeglichkeiten-werden-professioneller-205926