Digitalisierung / IT-Sicherheit II

Herausforderung Deepfake-Identitätsbetrug

Deepfake-Technologien erlauben es mittlerweile Betrügern, mit geringem Aufwand immer bessere Identitätsfälschungen zu begehen. Betroffen sind längst nicht mehr nur Personen des öffentlichen Lebens. Das Phänomen wird zunehmend auch für Unternehmen und ihre Kunden zum Problem.

› Michael Born

Deepfakes sind mit künstlicher Intelligenz (KI) hergestellte Bilder oder Videos, die authentisch wirken, es aber nicht sind. Sie lassen sich mittlerweile in erschreckend guter Qualität selbst von nicht IT-affinen Menschen mit geringen Programmierkenntnissen herstellen.

Während Deepfakes in der breiten Öffentlichkeit hauptsächlich im Zusammenhang mit Desinformationskampagnen in der Politik oder mit Verunglimpfung von Prominenten, Rachepornos oder Kinderpornografie in Zusammenhang gebracht werden, nutzen Kriminelle die Methode immer häufiger für Identitätsdiebstahl als Mittel für finanzielle Betrügereien.

KI-gestützte Betrugsversuche gehören mittlerweile denn auch zu den häufigsten Betrugsformen. Experten gehen davon aus, dass sich die Zahl der weltweit identifizierten Deepfakes zwischen 2022 und 2023 verzehnfacht hat. Dabei ist die Identitätskarte das am häufigsten betroffene Identitätsdokument. Und es sind bei weitem auch nicht mehr nur Identitätsdokumente von Ländern, deren Ausweis eine geringe Fälschungssicherheit bietet, betroffen. Aufgrund dieser rasanten Entwicklung sind Deepfakes längst nicht mehr nur für Prominente ein Problem. Sie stellen mehr und mehr auch für Firmen, die sich bei ihrem Geschäft auf digitale Identitätsprüfung und KYC-Prozesse (Know Your Customer) verlassen, eine Herausforderung dar.

kurz & bündig

- Deepfakes sind mit künstlicher Intelligenz (KI) hergestellte Bilder oder Videos, die authentisch wirken, es aber nicht sind. Kriminelle nutzen die Methode immer häufiger für Identitätsdiebstahl als Mittel für finanzielle Betrügereien.
- Deepfakes stellen mehr und mehr auch für Firmen, die sich bei ihrem Geschäft auf digitale Identitätsprüfung und KYC-Prozesse (Know Your Customer) verlassen, eine Herausforderung dar.
- Ist die Echtheit von Bildern umstritten, kann dies zu Problemen bei der Gewährung von Persönlichkeits-, Eigentums- und anderen Grundrechten bis hin zu Justizirrtümern führen.

Grosses Schadenpotenzial

Digitale Identitätsverifikationslösungen sind zwar bereits in der Lage, mit diversen Betrugsarten umzugehen. So können beispielsweise Deepfakes, bei denen nicht physisch vorliegende ID-Dokumente oder gefälschte Selfies verwendet werden, erkannt werden. Dazu kommen genauso wie bei der Herstellung von Deepfakes Methoden der generativen KI respektive des Machine Learnings, speziell des Deep Learnings, zum Einsatz. Es ist aber im Zuge der Fortschritte im Bereich KI damit zu rechnen, dass die Zahl der Fälschungsangriffe, mit denen die heute verwendeten Algorithmen nicht mehr umgehen können, zunehmen werden.

Identitätsdiebstahl kann indes auf vielen Ebenen Schäden verursachen. Von unberechtigtem Zugang zu Bankkonten von Privatpersonen über Kreditbetrug bis hin zur Gefährdung von finanziellen Unternehmensressourcen. Mehr noch: Wenn Unternehmen und deren Kunden elek-

tronischen Identifikationsprozessen nicht mehr trauen und solche Verfahren von der Politik abgelehnt werden, müsste die Wirtschaft wieder auf rückständige, zeitaufwendige analoge Prozesse zurückgreifen – mit unliebsamen Folgen sowohl für Unternehmen als auch für Verbraucher. Dem Marktforschungsunternehmen Gartner zufolge soll aufgrund von Deepfakes, die durch KI erzeugt wurden, bis 2026 ein Drittel der Unternehmen biometrische Gesichtserkennung als unzuverlässig einstufen.

Verbesserung der Technologien

Aus den genannten Gründen ist es von eminenter Wichtigkeit, dass die Identitätsverifikationsindustrie weiter daran arbeitet, den Betrügern das Handwerk zu legen. Technologien zur Identitätsprüfung müssen kontinuierlich verbessert werden, um Unternehmen langfristig zu ermöglichen, sich vor KI-generierten Deepfakes zu schützen. Nun gibt es zwar auf dem Markt erste Software-Lösungen, die Deepfakes mit hoher Treffsicherheit entlarven können sollen. Allerdings funktionieren solche meist nur gut in kontrollierten, experimentellen Umgebungen, bei denen die den Deepfakes zugrunde liegenden Modelle bekannt sind.

Am vielversprechendsten sind für die automatisierte Erkennung von Deepfake-Identitätsbetrug nach heutiger Kenntnis Methoden, die wie bei der Erstellung von Deepfakes auf Deep Learning setzen. PXL Vision arbeitet unter anderem zusammen mit dem Forschungsinstitut Idiap an neuartigen Methoden, bei denen unter Verwendung von synthetischen Gesichtsbild-Datensätzen die Genauigkeit der Identitätsverifikation hinsichtlich Alter, Hautfarbe und Geschlecht verbessert werden soll (siehe Box).

Unternehmen gefordert

Wenn Unternehmen mit ihrem Geschäftsmodell auf digitale Identifikations- und

KYC-Prozesse angewiesen sind, sollten sie gewisse Handlungsempfehlungen beachten. Zumdem sollten sie die Regulierung zu KI und Deepfakes im Blick behalten.

Der Bundesrat hat Ende 2023 mögliche Regulierungsvorschläge für künstliche Intelligenz beim Bundesamt für Umwelt, Verkehr, Energie und Kommunikation UVEK in Auftrag gegeben. Ziel ist es, das Potenzial von KI nutzbar zu machen und gleichzeitig die Risiken für die Gesellschaft zu minimieren.

In der EU sollen mit dem geplanten «AI Act» betrügerische Anwendungen von Deepfakes gesetzlich geregelt werden. Man setzt dabei auf besondere Transparenzanforderungen und eine Kennzeichnungspflicht für mithilfe von KI manipulierten Inhalten. Das Gesetz könnte 2026 in Kraft treten. Es ist zu erwarten, dass sich die schweizerische Gesetzgebung an die der EU anlehnen wird.

Darüber hinaus sollten Firmen in die nötigen Kompetenzen investieren, um mit Deepfakes angemessen umgehen zu können. Insbesondere empfiehlt es sich, die Entwicklung von automatisierten Deepfake-Erkennungslösungen zu verfolgen und aktiv mit den Forschenden und Unternehmen zusammen zu arbeiten. Gemeinsam können Anwender- und Anbieterunternehmen neue Lösungen schneller zur Marktreife bringen.

Die Bedeutung des Vertrauens

Die stetige Verbesserung von automatisierter Deepfake-Erkennung ist aus Gründen der Cybersicherheit für verschiedene Branchen wichtig. Doch nicht nur für eine sichere digitale Wirtschaft hat die Authentizität von Bildern eine grosse Wichtigkeit. Ist das Vertrauen in die Echtheit von Bildern nicht vorhanden, gefährdet dies nämlich nicht nur digitale Identifikationsabläufe für Kunden-Onboarding oder andere Antragsprozesse. Ist die Echtheit von Bildern umstritten, kann dies zu Problemen bei der Gewährung von Persönlichkeits-, Eigentums- und anderen Grundrechten bis hin zu Justizirrtümern führen. Letztlich geht es um die Aufrechterhaltung der wirtschaftlichen und politischen Stabilität ganzer Gesellschaften. «

Innovationsförderung des Bundes

Mit künstlicher Intelligenz hergestellte Deepfake-Gesichtsbilder

Das Schweizer ETH-Spin-off PXL Vision beschäftigt sich schon lange mit den neuen Herausforderungen des Identitätsbetrugs mit Deepfakes. Die Screen-Detection- und Injection-Attack-Technologien von PXL Ident sind bereits heute in der Lage, eine Vielzahl von Deepfake-Versuchen zu erkennen. Seit Februar 2024 arbeitet das Unternehmen zudem mit dem renommierten Idiap Research Institute an der Entwicklung einer KI-basierten Lösung zur Erkennung von Deepfake-Betrug bei Gesichtsbildern und Reisedokumenten.

Ziel des von der Schweizerischen Agentur für Innovationsförderung Innosuisse unterstützten Projekts ist es, bis Ende 2025 eine weltweit führende KI-basierte Technologie zur Erkennung von Gesichtern und Reisedokumenten zu entwickeln. Diese Technologie soll die Sicherheitsstandards von digitalen Identitätsprüfungslösungen weiter erhöhen.

Quelle (Foto): Idiap Research Institute

Porträt

Michael Born
CEO, PXL Vision

Michael Born ist CEO des auf digitale Identitätsprüfung spezialisierten ETH-Spin-offs PXL Vision. Es unterstützt Unternehmen dabei, das digitale Kunden-Onboarding zu automatisieren, Konversionsraten zu steigern, die Kosten für Onboarding und Compliance zu senken und Identitätsbetrug zu verhindern. Basierend auf künstlicher Intelligenz bietet PXL Vision vollautomatische, webbasierte ID-Verifikationslösungen, die individuell konfiguriert und nahtlos in bestehende Prozesse integriert werden können.

Kontakt

michael.born@pxl-vision.com
www.pxl-vision.com