Die Schattenseite der künstlichen Intelligenz

Deepfakes als Milliardenrisiko für die Versicherungswirtschaft. Der erste bekannte Versicherungsfall unter Nutzung künstlicher Intelligenz hat den Versicherer Euler Hermes bereits 220.000 € gekostet.

Der CEO-Fraud oder auch Fake President Betrug hat unzähligen Unternehmen hohe Geldbeträge gekostet. Der wohl bekannteste Fall betrifft einen Flugzeugkomponentenhersteller, der durch Betrüger einen zweistelligen Millionenbetrag verloren hat. Durch interne Kontrollsysteme, insbesondere die stringente Durchsetzung eines 4-Augen-Prinzips und regelmäßige Sensibilisierungsschulungen der Mitarbeitenden, hat der CEO-Fraud für die Versicherungswirtschaft an Bedeutung verloren. Payment Fraud und das Fehlleiten von Waren sorgen seit geraumer Zeit für Schadenfrequenz bei den Versicherern. Doch auch auf dieses Betrugsszenario stellen sich Unternehmen jetzt ein.

Nachdem bereits unterschiedliche Varianten des CEO-Fraud bekannt geworden sind – beispielsweise durch die Einbindung eines angeblichen IT-Mitarbeitenden, der in Zusammenarbeit mit der Polizei eine Scheinüberweisung zur Überführung der Täter anweist – ist bereits ein erster Schadenfall bekannt geworden, bei dem Täter künstliche Intelligenz genutzt haben. Konkret kam eine Software zur Imitation der Stimme zum Einsatz.

Was ist passiert? Die Täter sind nach dem klassischen CEO-Fraud Tatmuster vorgegangen. Der Unterschied lag lediglich darin, dass der falsche Chef plötzlich die richtige Stimme hat. Dies macht eine Enttarnung der Täter ungleich schwerer. Neben den bekannten Zahlungsanweisungen per E-Mail gab der vermeintliche Chef die Anweisung auch noch telefonisch. Diese telefonische Anweisung überwand das Misstrauen des Mitarbeitenden und dieser wies die Zahlung von 220.000 € an. Das Geld konnte nicht zurückgeholt werden.

Täter können den Betrug auf diese Weise beliebig fortentwickeln. Neben Stimmimitationssoftware können durch künstliche Intelligenz inzwischen auch täuschend echt wirkende Videos von Personen hergestellt werden, sogenannte Deepfakes. Die Versicherungswirtschaft sieht hierin ein erhebliches Risiko. Wer die Gestik oder Mimik und auch Stimmen von Vorstandsvorsitzenden oder in der Öffentlichkeit stehenden Personen nachahmen und sie willkürlich Sätze sagen lassen kann, hat starke Werkzeuge für Betrug und andere Formen der Desinformation in der Hand. Der Begriff der Desinformation existierte bereits lange vor dem Internet, das Ausmaß der Verbreitung und damit die Wirkung, die Desinformationskampagnen erzielen können, nimmt jedoch durch die weltweite Vernetzung in Form der fortschreitenden Digitalisierung dramatisch zu.

Die Herstellung eines Deepfake bedeutet nichts anderes als das Auswechseln des Gesichts einer Personen A gegen das Gesicht einer Person B. Dem Ergebnis liegen die Mimik, Gestik und Aussagen der Person A zugrunde, zu sehen ist jedoch das Gesicht von Person B. In technischer Hinsicht nutzt die Software eine simple Technik. Dieser liegt das sogenannte Deep Learning zugrunde. Hierbei kommen sogenannte Autoencoder zur Anwendung. Dabei handelt es sich um künstliche neuronale Netze, welche darauf spezialisiert sind, eingegebene Daten zu komprimieren und aus diesen komprimierten Daten wiederum ein möglichst realitätsgetreues Abbild des Originals herzustellen. Verkürzt dargestellt sind Deepfakes das Produkt zweier Algorithmen, die in einem Generative Adversarial Network (erzeugendes gegnerisches Netzwerk) zusammenarbeiten, hierbei werden neue Daten aus bestehenden Datensätzen generiert. Benötigt wird ausreichend Videomaterial zweier Personen. Je mehr Material vorliegt, desto besser werden die Ergebnisse, da dem Algorithmus mehr und in der Regel vielfältigere Lernvorlagen zur Verfügung stehen. Wer meint, von ihm existiere nicht ausreichend Bildmaterial für die Erstellung eines Deepfakes, der dürfte sich täuschen. Für manierliche Ergebnisse genügen ca. 300 Bilder und ein einsekündiges Video liefert bereits ca. 20 – 30 Bilder, eine Minute Videomaterial etwa 1.500 Bilder. Über Aufnahmen von öffentlichen Veranstaltungen und Social Media liefert das Internet auch für Privatpersonen umfangreiches Bildmaterial, welches vielfach frei verfügbar ist.

Heutzutage existiert Software (Open Source) zur Herstellung von Deepfakes, die für jeden Einzelnen frei zugänglich ist, die prominentesten Lösungen sind Face App und DeepFaceLab. Ein Selbstversuch mit dem heimischen Rechner zeigt, dass es binnen weniger Wochen auch Layen möglich ist, sehr manierliche Ergebnisse zu erzielen. Das Netz bietet insbesondere über Youtube umfassende Tutorials, mit denen das Handling der Software sich leicht erlernen lässt.

Im Ergebnis kann man mit Deepfakes jedwede Person Dinge sagen oder tun lassen, die sie nie gesagt oder getan hat. Bekannt gewordene Beispiele sind Deepfakes von Mark Zuckerberg, der in einem Video vorgibt, Millionen von Daten gestohlen zu haben oder dem vermeintlichen Barack Obama, der Donald Trump als Vollidioten betitelt. Erst bei sehr genauem Betrachten lassen sich diese Videos als Fake enttarnen.

So schnappte wohl auch die Fake-Falle zu, als der vermeintliche deutsche Vorstandschef eines Energieunternehmens bei seinem Pendant aus der britischen Niederlassung durchklingelte und Zahlungen anwies. Allianz-Tochter Euler Hermes gab letztes Jahr eine Pressemeldung dazu heraus, die medial vielfach aufgegriffen wurde.

Durch **Deepfakes wird eine ohnehin schon komplexe Risikowelt noch komplizierter. Unternehmen werden gezwungen sein, neue Kontrollmechanismen zu entwickeln, um auch Deepfakes enttarnen zu können. Beispiele hierfür können Trickfragen an einen Anrufer oder aber das Auffordern zum Durchführen ungewöhnlicher Bewegungen in Videokonferenzen sein. Weiterhin bleibt die stringente Einhaltung interner Kontrollmaßnahmen und eine offene Kommunikationskultur innerhalb des Unternehmens das wirksamste Instrument, um möglichen Betrugsschäden vorzubeugen.**

Bislang finden sich Deepfakes noch überwiegend im Bereich der Filmindustrie oder werden wie im Falle Mark Zuckerberg oder Barack Obama zur Belustigung genutzt. Erste Fälle versuchter Wahlmanipulationen durch Deepfakes gab es bereits in den USA und auch der Betrugsfall mittels Stimmimitation zeigt, dass die Täter künstliche Intelligenz zu nutzen beginnen, um ihre Betrugsmethoden weiter zu verbessern. Mag das Risiko heute noch überschaubar sein, sollte es dennoch nicht unterschätzt werden. Neben Betrug könnten Täter mittels Deepfakes auch Unternehmen mit der Veröffentlichung brisanter Videos erpressen. Keine schöne Zukunftsperspektive - nicht für die Gesellschaft, aber auch nicht für die Wirtschaft oder die Versicherer - stellt schließlich auch die Warnung vor einer Fake-Kultur dar, in der man nicht mehr einschätzen kann, worauf man sich verlassen kann und nicht mehr glauben kann, was man mit eigenen Augen sieht oder mit eigenen Ohren hört.

Foto: Cyber | Financial Lines bei Funk International Austria GmbH

Autor: Süleyman Yenier, BSC

Bereichsleitung Haftpflicht | Rechtsschutz |

Cyber | Financial Lines bei Funk International Austria GmbH