Deepfakes: Mittels künstlicher Intelligenz erschaffene Personen können für kriminelle Zwecke eingesetzt werden.

Drohen, manipulieren, täuschen

Deepfakes – technisch veränderte Videos, Fotos oder Texte – können unangenehme oder gefährliche Folgen haben für Personen, Unternehmen oder den Staat. Eine interministerielle Arbeitsgruppe erarbeitet einen Aktionsplan gegen Deepfakes.

Deepfake wird als Überbegriff für verschiedene Formen der audiovisuellen Manipulation verwendet. Deepfakes sind beispielsweise Videos, in denen Gesichter oder sonstige Elemente ausgetauscht werden und beispielsweise Stimmen, Mimik und Lippenbewegungen imitiert oder übernommen werden. Ohne technische Unterstützung ist die Täuschung kaum entdeckbar. Personen werden in Videos Aussagen in den Mund gelegt oder sie begehen scheinbar Handlungen, die nie stattgefunden haben. Die Entwicklungen der künstlichen Intelligenz (KI) und vor allem der Deepfakes schreiten rasant voran.

Seit den 1990er-Jahren werden digitale Bilder mittels spezieller Programme bearbeitet und verändert. Die breite Verfügbarkeit solcher Programme führte dazu, dass es heute besser nachvollziehbar ist, ob ein Bild echt ist oder verändert wurde. Dennoch blieb in der Gesellschaft der Eindruck, dass man Fotos nicht trauen kann. Videos galten bislang als eindeutige Beweise dessen, was man sieht und hört. Deepfakes stellen das infrage. Die neue Technologie bietet ungeahnte Möglichkeiten für Kriminelle und ist eine Herausforderung für die Sicherheitsbehörden Österreichs und Europas. Aus sicherheitspolitischer Sicht gibt es fünf Bereiche, die Deepfakes zur Herausforderung bzw. Bedrohung machen können.

Destabilisierung des Staates. Deepfakes sind bedenklich, weil sie zur Destabilisierung des Staates und der Gesellschaft eingesetzt werden können. So kann etwa ein Video, das ein Staatsoberhaupt oder ein Regierungsmitglied zeigt, das Dinge sagt, die zu Massendemonstrationen, Regierungssturz und Staatskrise führen. Auch denkbar ist, dass ein einziges, spektakuläres Video, das ausgesprochen realistisch wirkt, zu einer Kette von Reaktionen in anderen Bereichen führt, wie dem Zusammenbruch der Aktienmärkte oder der Beeinflussung von Wahlen. Besonderes Augenmerk wird man in Zukunft auf Desinformationskampagnen durch ausländische Akteure haben müssen. Es kann nicht ausgeschlossen werden, dass diese Technologie sowohl von anderen Staaten als auch von terroristischen Organisationen missbraucht werden.

Desinformation nimmt eine neue Dimension an. Deepfakes und Artificial Personas – mittels KI erschaffene Personen – können zur Spionage eingesetzt werden. 2019 wurde in den USA ein *LinkedIn*-Konto aufgedeckt, das „Katie Jones“, Forscherin eines führenden US-Think-Tanks, gehörte. Sie war

FOTO: SONO CREATIVE/STOCK.ADOBE.COM

jedoch eine künstlich geschaffene Person und Teil einer Spionageoperation. Zum Zeitpunkt der Entfernung durch *LinkedIn* 2019 hatte die Person bereits mit mehreren Mitarbeitern der US-Regierung Kontakt gehabt.

Zur Bedrohung für die innere Sicherheit können auch verfälschte Videos werden, die etwa Polizeigewalt zeigen oder Polizisten präsentieren, die strafbare Handlungen begehen. Die Konsequenzen können etwa der Verlust des Vertrauens in die Polizei oder Massendemonstrationen sein.

Wirtschaft. Die zweite Herausforderung stellt sich für die Wirtschaft. Die Europäische Agentur für Cyber-Sicherheit (ENISA) geht in ihrem Cyber-Bedrohungsbericht 2020 davon aus, dass Cyber-Kriminelle verstärkt zu Deepfakes greifen werden, um Unternehmen zu erpressen. Der CEO-Fraud (Geschäftsführerbetrug) oder Erpressungen im Allgemeinen werden mithilfe von Deepfake-Technologien eine neue Dimension annehmen.

Ein Video, das einen Mitarbeiter eines Unternehmens beim Begehen einer strafbaren Handlung zeigt, um diesen dann zu diskreditieren oder zu erpressen, ist denkbar. Die Drohung der Veröffentlichung eines Videos über ein Unternehmen könnte dazu führen, dass Aktienkurse abstürzen oder Geschäftspartner irritiert werden.

Gesellschaft und Medien. Drittens stehen Gesellschaft und Medien vor neuen Herausforderungen. Es kann zu einem kontinuierlichen Verlust des Vertrauens in digitale Inhalte kommen. Wenn eine große Menge an Deepfake-Videos mit politischem Inhalt hochgeladen wird, deren Verbreitung nicht mehr eingedämmt werden kann, kann dies zur Infragestellung staatlicher Institutionen führen oder der Beeinflussung von Wahlen.

Faktoren, die die Bedrohung durch Deepfakes beschleunigen, sind Videoplattformen oder Private-Messaging-Plattformen. Nach dem gleichen Prinzip wie Fake News spielt die Authentizität eines Videos oft keine Rolle, nachdem es mehrfach über Social Media geteilt wurde. Wenn sich Deepfakes als störende Elemente im Internet verbreiten, schwächen sie das Vertrauen der Bürger. Sie können zu einem Risiko für die innere Sicherheit werden.

Deepfakes: Gefälschtes Bildmaterial kann als digitaler Beweis für beliebige Situationen genutzt werden.

Bedrohungsszenarien für Personen. Viertens ergeben sich Bedrohungsszenarien für Personen. Begonnen hat das Thema Deepfakes 2018, als ein Nutzer der Plattform *Reddit* auf die Idee kam, Bilder von Frauen aus dem Netz zu nehmen und mit KI-basierter Technologie für pornografische Videos zu nutzen. Das Verwenden des Gesichts einer bestimmten (unbeteiligten) Person für die Produktion eines Fake-Pornofilms betrifft bis dato vor allem Frauen und Kinder.

Einem Bericht der Firma *Deeptrace* zufolge waren 2019 14.678 Deepfake-Videos online. Davon war die Verteilung der Videos zwischen pornografischen und nicht pornografischen Inhalts 96 zu 4 Prozent. 2019 wurde die App „Deep Nude“ geschaffen, die es ermöglichte, von einem hochgeladenen Foto einer Frau die Kleidung zu entfernen, um sie nackt zu zeigen. Die App ging rasch wieder offline, Nachahmung ist nicht ausgeschlossen.

Strafverfolgung. Zuletzt können sich in diesem Bereich neue Herausforderungen für die Strafverfolgung stellen. Denkbar ist ein Szenario, in dem ein Verdächtiger auf seinem Handy ein Video besitzt, das ihn z. B. in London zeigt, während er in Wien eine Straftat begeht. Straftäter könnten die Technologie einsetzen, um Überwachungsvideos zu manipulieren. Oftmals sind Videoaufnahmen von Verdächtigen wichtige Belege. Gelingt es den Tätern, die Überwachungsvideos zu manipulieren, werden die Ermittlungen erschwert. Gefälschtes Bildmaterial kann als digitaler Beweis für beliebige Situationen genutzt werden. Sobald sich Videos mittels KI-Verfahren in einer Weise verändern lassen, dass Manipulationen mit bloßem Auge nicht mehr zu erkennen sind, bringt das eine neue Herausforderung für die Strafverfolgung.

Aktionsplan gegen Deepfakes. Aufgrund der vielfältigen Auswirkungen von Deepfakes hat der Nationalrat im Oktober 2020 eine Entschließung angenommen, die die Bundesregierung auffordert, sich dem Phänomen der Deepfakes anzunehmen. Unter Federführung des Bundesministeriums für Inneres wird derzeit in einer interministeriellen Arbeitsgruppe ein Aktionsplan zum Thema Deepfakes ausgearbeitet. Das BMI und andere Ressorts arbeiten zusammen mit dem *Austrian Institute of Technology* an einem Sicherheitsforschungsprojekt zur Erkennung von Deepfakes. Zudem beteiligt sich Österreich an europäischen Initiativen (EU-Aktionsplan gegen Desinformation, European Democracy Action Plan, EU-Schnellwarnsystem gegen Desinformation).

Deepfake-Technologie bringt eine Menge an Herausforderungen, deren sich Politik, Strafverfolgung, Wirtschaft, Wissenschaft, Medien und jeder Bürger bewusst sein müssen, um damit vernünftig umgehen zu können. Man darf Deepfakes nicht nur als Bedrohung sehen, sondern muss auch den Einsatz der Technologie in Bereichen wie Wissenschaft, Kunst oder Bildung ermöglichen und würdigen. Deepfakes können zu legitimen Zwecken verwendet werden, wie die politische Satire. Eine klare Differenzierung der legitimen Anwendungsbereiche und eine frühzeitige Auseinandersetzung mit dem Thema sind notwendig. *Caroline Schmidt*

FOTO: NSATA_PRODUCTION/STOCK.ADOBE.COM