Deepfakes und mit KI manipulierte Realitäten: Analyse für die Schweiz

Der Textinhalt stammt von Frau Laetitia Ramelet; sie ist Projektleiterin bei TA-SWISS, der Schweizer Stiftung für Technologiefolgen-Abschätzung.

Der technologische Fortschritt hat unser Leben in vielen Bereichen erleichtert, z.B. durch vereinfachte Kommunikation und schnellen Zugang zu Informationen. Doch wo Licht ist, ist auch Schatten. Die heutige Digitalisierung bietet auch eine Plattform für betrügerische Handlungen. Ein besonders relevantes Beispiel sind die sogenannten „Deepfakes“. Was Deepfakes genau sind und welche Chancen und Risiken sie mit sich bringen, wird in diesem Factsheet erläutert.

«**Deepfakes**» sind mithilfe von KI-Techniken erstellte Bild-, Video- oder Toninhalte, die authentisch erscheinen. Wenn in diesen Inhalten eine Person etwas tut oder sagt, was sie nie getan oder gesagt hat, sind die Gefahren spürbar: Identitätsdiebstahl, Rufschädigung, Verbreitung von Falschinformationen und pornografische Inszenierungen ohne Einwilligung der betroffenen Person. Für die Ausbildung, die Unterhaltung, die Kunst oder die Werbung bieten dieselben Technologien jedoch Chancen; dies mit der Gestaltung von vollständig neuem Ton-, Bild- und Videomaterial.

Die Studie von TA-SWISS liefert fundiertes Wissen zu den Chancen und Risiken dieser Technologien.

Politische Einflussnahme und Meinungsäusserungsfreiheit

Deepfakes können ein Mittel der Einflussnahme auf die Meinungsbildung der Bevölkerung und auf die **politischen Abläufe in einer Demokratie** sein. Problematisch ist es, wenn mit diesen Technologien politische Persönlichkeiten, ihre politische Haltung oder ihre Partei diskreditiert oder Falschinformationen verbreitet werden. Es ist auch denkbar, dass mit falschen Gesichtern und Profilen im Internet Bürgeraktionen simuliert werden könnten. Gleichzeitig eignen sich Deepfakes u.a. für humoristische oder satirische Zwecke ohne manipulatorische Absicht. Sie sind Teil der freien Meinungsäusserung, die ein zentraler Pfeiler der Demokratie darstellt.

Zwischen wirtschaftlicher Innovation... und Angriffen gegen Unternehmen

Für die **Wirtschaft** gibt es eine Vielzahl an vielversprechenden Anwendungen von Deepfakes und synthetischen Medien. Mit ihnen können Lieder, Filmszenen und Videospiele ebenso wie neue visuelle und interaktive Formate für die Kommunikation und Werbung erstellt werden. Doch auch für Unternehmen kristallisieren sich Risiken heraus. Durch die widerrechtliche Aneignung der Identität einer Person können Mechanismen zur Identitätsprüfung getäuscht, sensible Informationen gewonnen und die Auszahlung von Geld oder Zugang zu einem System für einen Cyberangriff erhalten werden.

Bevölkerungsumfrage: Deepfakes zu erkennen ist schon heute schwierig

Ein von TA-SWISS durchgeführtes Onlineexperiment zeigt auf, dass die befragten Personen ein technisch gut gemachtes Deepfake-Video praktisch nicht von einem realen Video unterscheiden können. Dies auch, wenn sie im Voraus Tipps zum **Erkennen von Deepfakes** erhalten. Laut der im September 2023 durchgeführten Umfrage kannten nur 57 Prozent der Befragten den Begriff «Deepfake» und 49 Prozent gaben an, bereits einen gesehen zu haben. Nur zwei Prozent hatten bereits einen Deepfake selbst erstellt, und drei Prozent bereits einen geteilt.

Rechtsgrundlagen existieren, sind aber oft nur schwer durchsetzbar

Das **Recht** muss einerseits dem Missbrauch von Deepfakes vorbeugen und andererseits deren legale Verwendung als Mittel der freien Meinungsäusserung schützen. In der Schweiz werden Deepfakes bereits in den rechtlichen Grundlagen des Persönlichkeitsschutzes (z. B. Bildrechte) und des Datenschutzes sowie im Strafrecht (z. B. im Fall der üblen Nachrede und des Identitätsmissbrauchs) erfasst. Die Anwendung dieser Bestimmungen bleibt aber beschränkt, da sich die grossen Onlineplattformen, auf denen Deepfakes kursieren, im Ausland befinden und die Verantwortlichen für Vergehen nicht immer identifizierbar sind.

Kein technisches Wundermittel

Zur Prävention kann auf die **Authentifizierung** von originalen Inhalten (mit digitaler Signatur) gesetzt werden, ebenso wie auf die entsprechende Kennzeichnung der von einer KI erstellten Inhalte. Es gibt zudem Software für die **Erkennung** von Deepfakes. All diese Vorgehensweisen sind nützlich, aber fehleranfällig und können technisch überwunden werden. Gleichzeitig werden die Manipulationstechniken immer besser.

Medien in der Pflicht

Die von TA-SWISS mit **Medienredaktionen** geführten Gespräche weisen darauf hin, dass Deepfakes von den Medienschaffenden in erster Linie als eine Form der Desinformation wahrgenommen werden. Folglich seien die Grundsätze des Journalismus – wie die Überprüfung der Quellen – das beste Mittel, um Deepfakes zu enthüllen. Zusätzlich haben mehrere Redaktionen für technisch komplexe Fälle Spezialteams gebildet, die digitale Inhalte prüfen.

Einige Empfehlungen der TA-SWISS-Studie

- **Onlineplattformen regulieren**, auf denen die meisten Deepfakes geteilt werden: Zusammenarbeit mit den Strafverfolgungsbehörden, Meldesystem für rechtswidrige Inhalte, Sperrung gemeldeter Deepfakes im Fall des Verdachts auf einen Verstoss und gleichzeitig Transparenzvorgaben und Widerspruchsmöglichkeiten im Falle unberechtigter Löschungen.
- **Bildung und Selbstverantwortung der Bürgerinnen und Bürger fördern:** Förderung der Medienkompetenzen, insbesondere im Umgang mit Social Media (z. B. Quellenprüfung und kritische Herangehensweise an im Internet zugängliche Inhalte) und Sensibilisierung, um das Teilen illegaler Online- Inhalte zu verhindern.
- **Beratungszentren für Opfer von Cyberkriminalität unterstützen**, da es den Opfern oft an Ressourcen fehlt, um auf einen Deepfake-Angriff wirksam zu reagieren.
- **Schweizer Unternehmen und Institutionen für den Umgang mit Deepfakes sensibilisieren**: interne Beurteilung der Risiken und Präventionsmassnahmen (z. B. Weiterbildung, Mechanismus für den Umgang mit allfälligen schädlichen Deepfakes, fortschrittliche Authentifizierungsmassnahmen).
- **Journalistische Standards hochhalten,** sowohl bei der Ausbildung der Medienschaffenden als auch durch eine Unterstützung des Schweizer Presserats, da die Medien beim Erkennen von Deepfakes und bei der Information der Bevölkerung eine wichtige Rolle spielen.
- **Ausarbeitung internationaler Normen gegen die Cyberkriminalität unterstützen**, da dieser nicht einzig auf nationaler Ebene beizukommen ist.

Die Studie und ihre Kurzfassung können auf der Projektseite von TA-SWISS heruntergeladen werden.

Konsumentenforum in Zusammenarbeit mit TA-SWISS Stiftung für Technologiefolgen-Abschätzung