Innovation «made in Switzerland»

# Erkennung von Deepfakes

**Ein Beispiel eines Unternehmens, das auf der «Swiss Cybersecurity Start-up Map» gelistet ist, ist PXL Vision. Zusammen mit dem Schweizer Forschungsinstitut Idiap entwickelt es eine Lösung zur Erkennung von Deepfakes und wird dabei von Innosuisse unterstützt.**

**Thomas Berner**

KI-generierte Betrugsversuche werden immer ausgefeilter und machen auch vor offensichtlich gefälschten Bildern und Videos nicht halt. Über sogenannte Deepfakes wird aber nicht nur betrogen, sondern auch politisch manipuliert. So werden schon mal einer linken Politikerin rechtspopulistische Worte in den Mund gelegt. Oder – vor allem weibliche – Film- oder Popstars finden sich plötzlich in pornografischen Darstellungen wieder. Doch auch abseits von solchen medienwirksamen Fällen finden ausgefeilte Betrugsversuche über gefälschte Identitäten statt. «Identitätsfälschung mithilfe künstlich generierter Daten wird zu einem immer grösseren Problem bei der Bekämpfung von Betrugsversuchen im Internet. Gemeinsam mit Idiap arbeiten wir deshalb mit Hochdruck an der nächsten Generation von Deepfake-Erkennung, um unsere Lösungen zur digitalen Identitätsverifikation an die neuen technologischen Herausforderungen anzupassen und noch sicherer zu gestalten», sagt Michael Born, CEO bei PXL Vision. Dieses 2017 gegründete Unternehmen ist eine führende Anbieterin im Bereich der digitalen Identitätsprüfung und zählt namhafte Unternehmen aus der Finanz-, Versicherungs- und Gesundheitsbranche zu seinen Kunden.

## Fälschern einen Schritt voraus

Deepfake-Erkennung stellt selbst für die modernste Technologie nach wie vor eine technische Herausforderung dar. Um ein Video zu fälschen, werden Inhalte aus

© PXL Vision

*Echt oder falsch? Eine neue Software soll die Deepfake-Erkennung erleichtern.*

anderen Videos reingeschnitten. «Aktuell sind gefakte Videos häufig noch von blossem Auge als solche erkennbar», hält Michael Born fest und fährt fort: «Die Technologien für Bildmanipulationen werden aber immer ausgefeilter, auch getrieben durch generative KI. Somit muss auch die Erkennungssoftware immer besser werden. Unser Ziel ist, den Fälschern einen Schritt voraus zu sein.»

Wie genau die neue Lösung funktionieren wird, verrät Michael Born natürlich nicht im Detail. Aber im Wesentlichen werden Technologien entwickelt, die in der Lage sein sollen, mit heute noch unbekannten Angriffsarten umzugehen. Darüber hinaus soll die Genauigkeit der Gesichts- und Altersverifikation weiter verbessert werden, u.a. durch die Verwendung synthetischer Datensätze, die in Bezug auf Alter, Hautfarbe und Geschlecht ausgewogen sind. Die Unterstützung durch Innosuisse ermöglicht die Finanzierung eines Entwicklerteams von fünf bis zehn Personen. Das Vorhaben ist auf 18 Monate ausgelegt. Wie schnell darf also mit einer marktreifen Lösung gerechnet werden? Dazu Michael Born: «Erste neue Erkenntnisse werden wir wohl schon nach sechs Monaten in unsere bestehenden Lösungen einbauen können, weitere 15 Monate werden wir nutzen, um die Lösung weiter zu verbessern. Insgesamt rechnen wir mit 12 bis 24 Monaten, bis die Lösung weiter ausgereift sein wird.»

## Sichere digitale Identitätsverifikation

Die Kunden werden also bald profitieren können. «Die Entwicklungsresultate fliessen sofort in die bereits eingesetzten Lösungen ein. Für andere Anbieter der Identifikationsbranche können wir die Lösung auch lizenzieren», so Michael Born. Hat man auch neue Kundensegmente im Visier? Michael Born bejaht diese Frage. «Deepfakes sind ein branchenübergreifendes Problem. So könnten dereinst auch Medienunternehmen, Social-Media-Provider, Regierungsorganisationen, wie z.B. Geheimdienste, oder Bild- und Videoplattformen für die Sicherstellung der Intellectual Property zu unseren Kunden gehören.» Wird man da auch mit dem Standard «jpeg Trust», dessen Einführung für dieses Jahr geplant ist, arbeiten? Michael Born relativiert: «Bei diesem Standard geht es primär um die Rückverfolgbarkeit eines Bildes und den Nachweis, wer welche Veränderungen daran vorgenommen hat. Das ist insgesamt ein guter Ansatz. Grosse Anbieter wie Google werden diesem Standard wohl folgen, doch viele andere nicht. Ob sich vor diesem Hintergrund jpeg Trust durchsetzen wird, wird sich zeigen.» PXL Vision setzt deshalb andere Prioritäten und will gemeinsam mit Idiap an der nächsten Generation von Deepfake-Erkennung arbeiten, um ihre Lösungen zur digitalen Identitätsverifikation an die neuen technologischen Herausforderungen anzupassen und noch sicherer zu gestalten. ■