Pressemitteilung

Wien, 19.03.2024

KAMPF GEGEN DEEPFAKES: NEUES PROJEKT ZUR ERKENNUNG VON BILD- UND VIDEOMANIPULATIONEN

Neues KIRAS-Projekt – geleitet vom AIT Austrian Institute of Technology – fokussiert auf die Verbesserung der Erkennung und Bekämpfung von Deepfakes in digitalen Bild- und Videoinhalten

Ob DALL-E 3, Gesichtsfilter auf TikTok und Instagram oder DeepFaceLab: Technologien, die die Manipulation von Videos und Bildern ermöglichen, boomen, sind zunehmend einfacher zugänglich und markieren für viele Menschen einen Wendepunkt, wenn es um das Vertrauen in digitale Inhalte geht. Wie erkennen, was echt ist, wenn die Fakes immer realistischer sind? Von Behörden und Verwaltung über Medienorganisationen und Privatwirtschaft bis hin zur Zivilgesellschaft sind alle mit den Herausforderungen und Gefahren durch Deepfakes konfrontiert.

Fokus auf Tool-Entwicklung und Präventionsarbeit

Das von Spezialist:innen für Künstliche Intelligenz (KI) sowie Bild- und Videoanalyse am AIT koordinierte Projekt „defame Fakes“, finanziert im Sicherheitsforschungs-Förderprogramm KIRAS des Bundesministeriums für Finanzen, dreht sich um die Erforschung und Entwicklung geeigneter und effektiver Tools für die unterstützende semi-automatisierte Erkennung von Deepfakes in großen Datenätzen. Ebenso sollen präventive Awareness-Maßnahmen einen gesamtgesellschaftlichen Diskurs anstoßen und das Problembewusstsein schärfen. Ziel ist es, der kontinuierlichen Aushöhlung des Vertrauens in digitale Inhalte entgegenzuwirken und durch die Schaffung neuer technologischer Möglichkeiten zur Erkennung von Bild- und Videomanipulationen Unternehmen und die Gesellschaft gegen Manipulationen zu schützen, Bewusstsein im Umgang mit digitalen Informationen zu schaffen und dadurch das Vertrauen in digitale Medien zu stärken. Die Bedarfsträger des Projekts sind das Bundesministerium für Inneres (BMI) und das Bundesministerium für Landesverteidigung (BMLV).

Desinformation im Kontext hybrider Bedrohungen

„Die Demokratie und demokratische Meinungsbildung werden derzeit in einem noch nie dagewesenen Ausmaß in Frage gestellt. Durch KI können irreführende Inhalte schnell erstellt und mit großer Reichweite verbreitet werden“, erklärt Michael Suker, Leiter des Cyber Dokumentations- & Forschungszentrums/ZentDok an der Landesverteidigungsakademie. Es drohe die Gefahr, dass Deepfakes unbemerkt von der Bevölkerung zur Beeinflussung der öffentlichen Meinung rezipiert werden. Deepfakes können Teil von hybriden Bedrohungen sein, die darauf abzielen, die politische, wirtschaftliche und soziale Struktur eines Staates oder einer Organisation zu untergraben oder zu schädigen, indem verschiedene Mittel kombiniert werden. Dazu gehören Cyberangriffe, Desinformation und wirtschaftlicher Druck. Suker weiter: „Die Erkennung und

Bekämpfung von Deepfakes sowie die Verbesserung der Medienkompetenz sind daher von entscheidender Bedeutung, um die Integrität von Informationen zu schützen und die nationale und internationale Sicherheit zu gewährleisten."

Analyse sozialer, ethischer und rechtlicher Folgen

In allen Forschungsaktivitäten spielt die Analyse sozialer, ethischer und rechtlicher Implikationen der Verbreitung von Deepfakes, aber auch der Nutzung von Detektionstools eine zentrale Rolle: Was bedeutet es für die Unternehmenssicherheit, wenn die Stimme und das Bild eines CEOs mittels Deepfake-Technologie imitiert werden können? Wie können Privatpersonen präventiv vor Betrügereien geschützt werden, die durch generative KI noch glaubwürdiger werden? Welche Folgen haben Desinformation, Manipulation von Meinungsbildungsprozessen oder die Bewerbung extremistischer Gruppierungen durch Deepfakes für die Demokratie? Und welcher Regulierungsbedarf ist notwendig, um nicht nur den Herausforderungen von Deepfakes zu begegnen, sondern auch die Nutzung von Detektionstools sicher zu gestalten?

Interdisziplinäres Team forscht zu Erkennung von Deep Fakes

Koordiniert wird das Projekt vom AIT, das heute ein etabliertes Kompetenzzentrum im Bereich KI-basierter Anwendungen rund um den Schutz vor Desinformationskampagnen ist – national und international. Die am AIT entwickelten innovativen Lösungen zielen auf den Schutz der Medienkonsument:innen und damit der Demokratie als Grundfeste unserer Gesellschaft ab. Martin Boyer, Projektleiter und Senior Researcher am AIT: „Als Koordinator war es für mich wichtig, ein breit aufgestelltes Konsortium zusammenzustellen, um alle notwendigen Kompetenzen innerhalb des Projektes abzudecken. Denn beim Thema Deepfakes haben wir es mit vielfältigen Herausforderungen zu tun, die nur interdisziplinär bewältigt werden können".

Hinter defame Fakes steht deshalb ein interdisziplinäres Team. „Um den Einfluss von Deepfakes auf gesellschaftliche Zusammenhänge zu verstehen, ist es zentral, die Zivilbevölkerung nicht aus den Augen zu verlieren: Präventionsaktivitäten und Awareness-Maßnahmen müssen dabei alle erreichen – von Jugendlichen hin zu älteren Personen", betont Louise Beltzung vom Österreichischen Institut für angewandte Telekommunikation (ÖIAT). Das ÖIAT wird die Implikationen von Deep Fakes in diesem Bereich untersuchen, in enger Kooperation mit PwC Austria, die einen Schwerpunkt auf das Bedrohungspotenzial für Unternehmen legen. „Deepfakes bergen ein ernstzunehmendes Risiko für Unternehmen, indem sie die Glaubwürdigkeit und Sicherheit von Geschäftskommunikation gefährden. Sie eröffnen neue Wege für Betrug, Erpressung und Rufschädigung, was die Notwendigkeit für effektive Maßnahmen zur Erkennung und Abwehr unterstreicht", betont Roland Pucher, Leiter PwC Cybersecurity & Innovation Lab.

Das Ziel ist es, nicht nur die Auswirkungen von Deepfakes besser zu verstehen, sondern auch geeignete Gegenmaßnahmen zu entwickeln, um die Gesellschaft vor den Gefahren der Bild- und Videomanipulation zu schützen und das Vertrauen in digitale Medien wiederherzustellen. Die Forschungsergebnisse sollen auch in der Praxis weiterhelfen – etwa bei Schulungen von Journalist:innen. „Es ist wichtig, dass wir uns nach dem Projekt defalsif-AI erneut in einem kompetenten Konsortium mit verschiedenen Erfahrungen und Interessen wiederfinden. Es freut

uns, unser Wissen aus dem redaktionellen Alltag einbringen und gemeinsam aktuellen Herausforderungen entgegnen zu können“, sagt Florian Schmidt, Leiter des Faktencheck-Teams der APA – Austria Presse Agentur.

„Das ist ein bedeutsamer Schritt in die richtige Richtung“ sagt Alexander Janda, Generalsekretär des KSÖ – Kompetenzzentrum Sicheres Österreich, Disseminationspartner im Projekt: „Das Thema Deepfake ist nicht mehr nur Zukunftsmusik, sondern direkt in der Mitte der Gesellschaft angekommen. Die Erkennung von manipuliertem Bild- und Videomaterial, aber auch die Sensibilisierung der Bevölkerung sind daher von größter Wichtigkeit.“

Über „defame Fakes“

„defame Fakes“ wird im Sicherheitsforschungs-Förderprogramm KIRAS des Bundesministeriums für Finanzen finanziert. In enger Kooperation mit den Bedarfsträgern des Projekts BMI und BMLV, sind folgende Partner:innen beteiligt:

- AIT Austrian Institute of Technology
- APA Austria Presse Agentur
- PwC Österreich
- KSÖ Kompetenzzentrum Sicheres Österreich
- ÖIAT Österreichisches Institut für Angewandte Telekommunikation

Weitere Informationen finden Sie auf der Projektwebsite unter www.defamefakes.at.

Rückfragehinweis:

Mag. (FH) Michael W. Mürling
Marketing and Communications
AIT Austrian Institute of Technology
Center for Digital Safety & Security
T +43 (0)664 235 17 47
michael.muerling@ait.ac.at I www.ait.ac.at

Mag. Michael H. Hlava
Head of Corporate and Marketing Communications
AIT Austrian Institute of Technology
T +43 (0)50550-4014
michael.hlava@ait.ac.at I www.ait.ac.at