

03.10.2022 – 10:03 Uhr

Forschungsprojekt zur Erkennung von Falschinformationen unter Beteiligung von APA und ORF erfolgreich abgeschlossen



Wien (ots) -

KI-gestütztes Werkzeug gibt mit wenigen Klicks Hinweise auf manipulierte digitale Inhalte und gezielte Desinformation – keine technischen Vorkenntnisse bei AnwenderInnen erforderlich

Zwei Jahre lang forschte ein Konsortium rund um das AIT Austrian Institute of Technology an einem leicht bedienbaren Werkzeug, das AnwenderInnen mit Hilfe von Künstlicher Intelligenz beim Erkennen von Falschinformationen im Internet unterstützt. Denn die enorme und stetig ansteigende Informationsflut im Internet bringt neue Gefahren wie z.B. Desinformationskampagnen mit sich, die in ihrer Fülle nicht mehr ausschließlich manuell erkannt werden können. Nun wurde das österreichische Forschungsprojekt „defalsif-AI“, das im Oktober 2020 im Rahmen des nationalen Sicherheitsforschungsprogramms KIRAS startete, erfolgreich abgeschlossen. Die entwickelte Lösung wurde offen konzipiert und ist damit einfach weiterzuentwickeln sowie in andere Systeme integrierbar.

Die stetige Zunahme von Falschinformationen, potenziert durch Social Media und andere Filterblasen, sowie die Entwicklung von wirksamen Strategien und Instrumenten dagegen, sind zentrale Herausforderungen unserer Zeit. Nicht „Wahrheitsfindung“, sondern der Schutz von Demokratie und demokratischen Institutionen lautete vor diesem Hintergrund das übergeordnete Projektziel von defalsif-AI. „Künstliche Intelligenz wirkt dabei als System, das den Menschen bei der Überprüfung und Verifizierung digitaler Inhalte unterstützt, ihm aber keine Entscheidungen abnimmt“, so Martin Boyer, Projektleiter und Senior Research Engineer am AIT Austrian Institute of Technology im Center for Digital Safety & Security.

Die Medienunternehmen APA - Austria Presse Agentur sowie der ORF waren neben Enlite AI, Research Institute, Donau Universität Krems und ministeriellen Bedarfsträgern aus der öffentlichen Verwaltung als Projektpartner an „defalsif-AI“ beteiligt.

Einfach bedienbar und transparent

Zum Überprüfen bzw. Verifizieren von digitalen, multimedialen Inhalten (Videos, Fotos, Texte, etc.) lädt man diese auf einer Online-Plattform hoch oder gibt dort die entsprechende Webadresse ein, ergänzt Hintergrundinformation und drückt den Startknopf. Zusatzinformationen wie „Audio“ oder „Porträtbild“ helfen dem System, die richtigen Analysemethoden und -tools auszuwählen. Das System sucht dann auf technischer Ebene nach Anhaltspunkten für Desinformation oder manipulierte Inhalte. Im Anschluss erhalten AnwenderInnen die Ergebnisse einfach und verständlich dargestellt.

Das Werkzeug besteht derzeit aus mehreren Bausteinen, die modulartig erweiterbar bzw. adaptierbar sind. „Wir haben das Tool zukunftssicher konzipiert, um laufend weitere Analysemethoden und Komponenten hinzufügen zu können und das Werkzeug auf diese Weise stets weiterzuentwickeln“, betont Boyer. An die Ergebnisse aus dem Projekt knüpfen die Forscher:innen bereits an: Schon fünf Nachfolgeprojekte stehen in den Startlöchern, bei denen der Fokus u.a. auf der weiteren Automatisierung und Weiterentwicklung der Analysemethoden liegen wird, womit das Tool künftig auch für einen breiteren Anwenderkreis zugänglich werden wird.

Weitere Informationen zum Projekt: www.defalsifai.at

KIRAS-Sicherheitsforschungsprogramm: www.kiras.at

Gefördert wird „defalsif-AI“ vom Bundesministerium für Finanzen und der Österreichischen Forschungsförderungsgesellschaft (FFG) im österreichischen Förderungsprogramm für Sicherheitsforschung KIRAS.

Pressekontakt:

APA - Austria Presse Agentur
Unternehmenskommunikation
Tel.: +43 1 360 60-5710
kommunikation@apa.at
www.apa.at

AIT Austrian Institute of Technology GmbH

Michael Mürling
Marketing and Communications
Center for Digital Safety & Security
Giefinggasse 4 | 1210 Vienna | Austria
T +43 50550-4126 | M +43 664 2351747 | F +43 50550-4150
michael.muerling@ait.ac.at | <https://www.ait.ac.at>

Medieninhalte

BILD zu OTS - KI-gestütztes Werkzeug gibt mit wenigen Klicks Hinweise auf manipulierte digitale Inhalte und gezielte Desinformation



Diese Meldung kann unter <https://www.presseportal.ch/de/pm/100000096/100895882> abgerufen werden.