defalsif-AI

Detektion von Falschinformation mittels Artificial Intelligence

Programm / Ausschreibung	KIRAS, Kooperative F&E-Projekte, KIRAS Kooperative F&E-Projekte 2019	**Status**	abgeschlossen
Projektstart	01.10.2020	**Projektende**	30.09.2022
Zeitraum	2020 - 2022	**Projektlaufzeit**	24 Monate
Keywords	Medienforensik; Fusion; Fake News; Desinformation; Künstliche Intelligenz;		

Projektbeschreibung

Der defalsif-AI Antrag befasst sich mit dem Problem der Desinformation – umgangssprachlich „Fake News“ – im Zusammenhang mit Angriffen auf kritische Infrastruktur. Als kritische Infrastruktur stehen hier die Demokratie bzw. das Vertrauen der Öffentlichkeit in die Demokratie im Vordergrund. Die behördlichen Akteure im angestrebten Projekt – BKA, BMLV und das BMEIA – sehen mit Beunruhigung auf potentielle, externe Angriffe auf staatliche Kernbereiche unserer Demokratie – wie z.B. zielgerichtete Kampagnen in sozialen Medien, um Wahlen zu manipulieren. Analyst/innenen in diesen Ministerien benötigen neue Ansätze im Umgang mit groß angelegten Desinformationskampagnen auf operativer, strategischer und politischer Ebene. Die im Projekt vertretenen Medienorganisationen – APA und ORF – sind Kraft ihres Grundauftrags dazu verpflichtet, den Bürgerinnen und Bürgern zuverlässige, überprüfte und überprüfbare Nachrichten zur Verfügung zu stellen, um durch ihre Informations- und Meinungsbildungsfunktion eine Demokratie zu ermöglichen. Die wachsende Menge an digitalen Inhalten führt dazu, dass diese Organisationen nach verbesserten Verifikationsmethoden und -werkzeugen suchen. Gemäß diesen Anforderungen konzentriert sich der Antrag auf Forschung in den Bereichen audiovisuelle Medienforensik, Textanalyse und deren multimodale Fusion mittels Methoden der Künstlichen Intelligenz (KI) und des Maschinellen Lernens. Ein Schwerpunkt ist dabei auch die nachvollzieh- und interpretierbare Präsentation der Ergebnisse für technische Laien.
Primäres Projektergebnis ist die Demonstration eines Proof-of-Concept (PoC) für die Analyse von digitalen Inhalten im Internet, welcher eine erste Beurteilung von Text, Bild, Video und Audio auf Glaubwürdigkeit/Authentizität ermöglicht und so Grundlagen für weitere Handlungsempfehlungen schafft. Dabei werden Screening und Monitoring Tools – welche Themen, Trends, Häufungen oder Anomalien bei der Informationsverbreitung im Internet bei staatlichen Kernprozessen erfassen – miteinbezogen. Einzelne Medienobjekte, aber auch ganze Webseiten des Surface Web (z.B. Nachrichtenseiten) und der sozialen Medien (z.B. Twitter) werden betrachtet. Die Implementierung wird einer interdisziplinären Evaluierung bei den projektbeteiligten Bedarfsträger/innen unterworfen. Weitere Forschung wird sich auf die Bereitstellung und Erzeugung multimodaler Trainings- und Testdaten konzentrieren. Von rechtlich, geistes- und sozialwissenschaftlicher Seite erfolgt eine umfassende Analyse der praktischen Bedarfsanforderungen sowie eine Abschätzung potentieller Risiken und gesellschaftspolitischer Implikationen. Schließlich zeigt ein im Projekt erarbeiteter Verwertungsplan, wie der resultierende PoC weiterentwickelt und als produktives System sowohl in staatlichen als auch in privaten Einrichtungen eingesetzt werden

könnte.

Abstract

The defalsif-AI proposal addresses the problem of disinformation – colloquially known as “fake news” – in particular disinformation considered in the context of an attack on critical infrastructure, where the critical infrastructure in question is democracy and the public trust in democracy and its institutions itself. The official stakeholders in the proposal – the BKA, BMLV and the BMEIA – are particularly concerned about potential external attacks on the Austrian democratic process, for example through engineered social media attacks that attempt to manipulate the electoral process. Analysts in these ministries require new approaches for identifying and dealing with large-scale disinformation campaigns at the operative, strategic and political level. At the same time, the Austrian Broadcasting Corporation (ORF) and the Austria Press Agency (APA) are concerned about fulfilling their crucial roles as journalistic institutions, which includes to enable democracy through its information and opinion-forming function. These organizations are thereby also forced to seek improved methods and tools for evaluating ever-increasing volumes of digital media in terms of identification, verification and correction of sources. Based on these stakeholder requirements, the proposal focuses on research in the areas of audio-visual media forensics, text analysis and the multimodal fusion of these with the support of Artificial Intelligence (AI) and Machine Learning methods. A principal focus of this research will be in enhancing the comprehensibility and interpretability of the results for non-experts in the forensic/technical field.
The primary project result is a proof-of-concept (PoC) implementation that can operate on a variety of sources, including the surface web (e.g. news sites) and social media (e.g. Twitter). Screening and monitoring tools will be included to identify topics, trends, accumulations or anomalies in the dissemination of information on the internet regarding core government processes. The tool should be able to ingest either individual media objects or eventually entire web pages – analysing images, audio, video (e.g. so-called Deepfakes), and text material – providing the basis for recommended actions. This PoC will be demonstrated to and evaluated interdisciplinarily by the project stakeholders. Additional research within the project will focus on providing and generating multi-modal data necessary to train and test machine learning models. A legal and social science analysis and assessment will be carried out, recommending guidelines as well as the application-oriented derivation of technical and organizational measures for future compliant implementation and execution of disinformation analysis platforms. Finally, the project will deliver an exploitation plan detailing how the resulting proof-of-concept could be further developed and deployed as a productive system in both governmental and private agencies.

Projektkoordinator

- AIT Austrian Institute of Technology GmbH

Projektpartner

- Universität für Weiterbildung Krems
- Bundesministerium für europäische und internationale Angelegenheiten
- APA - Austria Presse Agentur eG
- Research Institute AG & Co KG
- EnliteAI GmbH
- Bundeskanzleramt
- Österreichischer Rundfunk
- Bundesministerium für Landesverteidigung