

FZI-Pressemitteilung

Mit Künstlicher Intelligenz gegen Desinformation: Projekt DeFaktS soll Fake News erkennen und bekämpfen

Die Zahl der online bewusst verbreiteten Falschmeldungen nimmt seit Jahren immer weiter zu. Ein interdisziplinäres Konsortium will mit dem Projekt DeFaktS dagegen vorgehen, indem es eine Künstliche Intelligenz (KI) so trainiert, dass sie Desinformation erkennen und über sie aufklären kann. Dazu nimmt das FZI mit drei Partnern verdächtige Social-Media- und Messenger-Gruppen unter die Lupe.

Karlsruhe, 12.04.2022 – Texte, Bilder und Videos: Der Ukraine-Krieg ist in den sozialen Medien sehr präsent. Unter der Vielzahl an Beiträgen befindet sich immer häufiger bewusst manipulativ verbreitete Desinformation – sogenannte Fake News – zum aktuellen Kriegsgeschehen. Genutzt wird dafür unter anderem Videomaterial aus Videospielen oder anderen Konflikten – zum Beispiel der Annexion der Krim im Jahr 2014 oder dem Syrien-Krieg 2020. Zuletzt war das Phänomen der Fake News beispielsweise auch im Zusammenhang mit der Corona-Pandemie und der Bundestagswahl 2021 zu beobachten. Mit der Verbreitung von Falschmeldungen („Fake News“) sollen Angst und Verunsicherung in der Bevölkerung geschürt und/oder das Vertrauen in die Politik erschüttert werden. Das Bundesministerium für Bildung und Forschung (BMBF) sieht in Desinformation eine zunehmende Gefahr für die Demokratie und den gesellschaftlichen Zusammenhalt. Daher fördert das BMBF derzeit im Forschungsrahmenprogramm der Bundesregierung zur IT-Sicherheit „Digital. Sicher. Souverän.“ mehrere Projekte, die Desinformation erforschen, um Gegenmaßnahmen zu entwickeln – darunter auch das Forschungsprojekt DeFaktS.

Mit Künstlicher Intelligenz gegen Fake News

DeFaktS – das steht für: „Desinformationskampagnen beheben durch Offenlegung der Faktoren und Stilmittel“. Das Projekt verfolgt einen umfassenden Ansatz zur Erforschung und Bekämpfung von Desinformation. Im Rahmen des Projekts extrahieren die Wissenschaftler*innen Nachrichten aus verdächtigen Social-Media- und Messenger-Gruppen und trainieren eine Künstliche Intelligenz (KI) so, dass die für Fake News charakteristische Faktoren und Stilmittel erkennen kann. „Typisch für Desinformation ist beispielsweise die Nutzung von emotionaler Sprache und Hyperbeln. Die KI wird nach dem Training in der Lage sein, derartige Indikatoren zu identifizieren“, erklärt Jonas Fegert, Abteilungsleiter am FZI.

Die trainierte KI wird im Anschluss als Komponente für eine XAI (Explainable Artificial Intelligence) genutzt. Eine XAI ist eine erklärbare Künstliche Intelligenz, die eindeutig nachvollziehbar macht, auf welche Weise die KI zu den Ergebnissen kommt. Auf Basis der XAI soll eine App entwickelt werden, die die User transparent über das mögliche Auftreten von Fake News informiert und warnt. Ein weiteres Ziel von DeFaktS: die XAI-Komponente über die Entwicklung einer Programmierschnittstelle auch für weitere Plattformen zugänglich machen. Auf diese Weise

FZI-Pressemitteilung

möchte das Konsortium zu einer Lösung beitragen, mit der sich Online-Plattformen möglichst automatisiert moderieren lassen.

Sensibilisieren, Vertrauen schaffen und Medienkompetenz stärken

Im Gegensatz zu bereits bestehenden Tools zur Detektion und Bekämpfung von Fake News soll die DeFaktS-App eine besonders einfach zu nutzende sowie universelle Lösung zur automatisierten Moderation von Online-Inhalten darstellen. Darüber hinaus identifiziert DeFaktS nicht nur für Fake News charakteristische Faktoren, sondern gibt den User*innen mithilfe des XAI-Ansatzes auch ein transparentes Feedback zum möglichen Auftreten von Fake News. Das bedeutet: Die Interpretation und Entscheidung, ob Fake News vorliegen oder nicht, liegt letztlich bei den Bürger*innen, die die Anwendungen nutzen. Es findet keine Filterung oder gar automatisierte Löschung von Inhalten statt. “Im Sinne der digitalen Bildung der Gesellschaft war uns dieser Aspekt sehr wichtig”, erklärt Jonas Fegert. “Wir möchten die Menschen sensibilisieren, sie über Strategien der Desinformationen aufklären und so Medienkompetenz stärken.”

Derzeit wird im Projekt mithilfe einer strukturierten Literaturrecherche eine Taxonomie von Desinformation erstellt. Das bedeutet, dass ein einheitliches Modell erarbeitet wird, mit dem sich Fake News nach bestimmten Kriterien klassifizieren lassen – dazu zählen Charakteristika, Dimensionen und Eigenschaften der Informationen.

Über das Projekt

Das Forschungsprojekt DeFaktS wird vom Bundesministerium für Bildung und Forschung (BMBF) mit rund 1.269.000 Euro über drei Jahre gefördert.

Das **FZI** übernimmt im Rahmen des Projekts die Rolle des Konsortialführers und ist maßgeblich an der Entwicklung der XAI-Komponente zur Detektion von Fake News sowie deren Evaluation beteiligt – unter anderem durch eine Laborstudie am Karlsruher Institut für Technologie. Außerdem erstellt das FZI Trainingsdaten mittels WebScraping, einem Verfahren, bei dem Texte ausgelesen werden, und führt Works hops sowie Experimente zur Erstellung der Datensätze durch.

Die **Murmuras GmbH** erstellt die Datensätze zur Klassifizierung der Fake News. Außerdem übernimmt das Unternehmen die Entwicklung und Erweiterung der DeFaktS-App-Architektur, die für die Feldstudie der Philipps-Universität Marburg zur Projektvalidierung benötigt wird.

Die **Philipps-Universität Marburg** entwickelt einen Algorithmus zur Datenextraktion von Texten und andere Datentypen. Ferner bereitet der Projektpartner die Datensätze auf und stellt diese als Application Programming Interface (API) – also als Programmierschnittstelle – den Partnern zur Verfügung. Die Universität ist zudem für die Reflektion und Validierung der entwickelten App zuständig und wird dazu eine Feldstudie anhand des Feedbacks der Netzenden durchführen.

FZI-Pressemitteilung

Liquid Democracy e.V. übernimmt im Projekt DeFaktS die Vorbereitung und Durchführung einer Anwendungsstudie, in welcher die entwickelte Anwendung in der realen Praxis erprobt und evaluiert wird.

Weitere Informationen über das Forschungsprojekt sind auf der [Projektseite](#) zu finden.

Über das FZI Forschungszentrum Informatik

Das FZI Forschungszentrum Informatik mit Hauptsitz in Karlsruhe und Außenstelle in Berlin ist eine gemeinnützige Einrichtung für Informatik-Anwendungsforschung und Technologietransfer. Es bringt die neuesten wissenschaftlichen Erkenntnisse der Informationstechnologie in Unternehmen und öffentliche Einrichtungen und qualifiziert junge Menschen für eine akademische und wirtschaftliche Karriere oder den Sprung in die Selbstständigkeit. Betreut von Professoren verschiedener Fakultäten entwickeln die Forschungsgruppen am FZI interdisziplinär für ihre Auftraggeber Konzepte, Software-, Hardware- und Systemlösungen und setzen die gefundenen Lösungen prototypisch um. Mit dem FZI House of Living Labs steht eine einzigartige Forschungsumgebung für die Anwendungsforschung bereit. Das FZI ist Innovationspartner des Karlsruher Instituts für Technologie (KIT).

Weitere Informationen

Daniela Bader, Communications
FZI Forschungszentrum Informatik
Haid-und-Neu-Str. 10-14, 76131 Karlsruhe
Telefon: +49 721 9654-979
E-Mail: presse@fzi.de
Internet: www.fzi.de