HYBRIS

Hybride Bedrohungs-Resilienz durch Interdisziplinäre Zusammenarbeit der Sicherheitsbehörden

Programm / Ausschreibung	KIRAS, Kooperative F&E-Projekte, KIRAS Kooperative F&E-Projekte 2021	**Status**	laufend
Projektstart	01.01.2023	**Projektende**	31.12.2024
Zeitraum	2023 - 2024	**Projektlaufzeit**	24 Monate
Keywords	Desinformation Erkennnung, Künstliche Intelligenz		

Projektbeschreibung

Obwohl Desinformation nun schon seit längerem ein anerkanntes Problem ist und aktiv auf nationaler und internationaler Ebene behandelt wird, gewinnt es tagtäglich weiter an Brisanz. Ausgehend von desinformationsgeleiteter Wählermanipulation, über mutwillige Beeinflussung der Pandemiebekämpfung, bis hin zu gezielter Kriegspropaganda. Letzteres hat auch die Aufmerksamkeit auf ein neues Phänomen gelenkt: Hybriden Bedrohungen. Diese abstrakte Zusammenfassung unterschiedlicher Bedrohungen durchkreuzt traditionelle Zuständigkeiten österreichischer Sicherheitsbehörden, in dem sie per Definition von staatlichen oder nicht-staatlichen Akteuren, durch Kombination offener und verdeckter militärischer und nichtmilitärischer Mittel durchgeführt. Dadurch fallen diese entweder in den Wirkungsbereich des BMI (z.B. Cyber-Security, Cyber-Crime), des BKA (z.B. Desinformation, öffentliche Sicherheit), oder des BMLV (z.B. Cyber-Defense). Ein konzertierter Angriff mittels Hybrider Bedrohungen erfordert deshalb die Zusammenarbeit der Sicherheitsbehörden, bzw. die Vereinfachung von Kooperationen oder den Austausch von Analyseergebnissen und Gegenstrategien. Dies ist der Ausgangspunkt, an dem HYBRIS ansetzt.

HYBRIS befasst sich mit hybriden Bedrohungen, welche die Absicht verfolgen, durch online koordinierte Operationen Überzeugungen und Einstellungen ausgewählter Zielgruppen zu beeinflussen, diese zum Handeln zu mobilisieren und in der Folge die physische und digitale Infrastruktur zu kompromittieren. Desinformation ist ein wesentlicher Bestandteil hybrider Bedrohungen und ein spezifischer Fokus im Projekt. Desinformationskampagnen zielen häufig darauf ab, Ängste in der Bevölkerung zu schüren, die unterschiedliche Folgen nach sich ziehen können. Unabhängig davon, ob es sich bei Nachrichten um Falschinformationen handelt oder nicht, ist es für Sicherheitsbehörden wichtig zu erkennen, ob von Reaktionen auf Nachrichten in Sozialen Medien und anderen Nachrichtenkanälen bzw. möglicherweise in der Folge organisierten Aktionen eine Bedrohung für Menschen oder kritische Infrastruktur ausgeht. Erst wenn es darum geht, geeignete Maßnahmen zu ergreifen, geht es im Falle der Verbreitung von Desinformation auch darum, durch die Erstellung und Verbreitung von auf Tatsachen und vertrauenswürdigen Informationen basierenden Gegendarstellungen einer bedrohlichen Entwicklung entgegenzuwirken.

Ein schnelles Reagieren auf diese Art von Nachrichtentrends ist besonders wichtig. Daher muss die inhaltliche Vorbereitung

möglicher Gegendarstellungen bereits beginnen, wenn diese noch im Entstehen sind. Eine der wesentlichen Herausforderung ist dabei die Gewährleistung eines umfassenden Überblicks über die aktuell vorherrschenden Trends. Die Schwierigkeit dabei ist, dass Informationen über eine unüberschaubare Zahl von Informationskanälen in hoher Geschwindigkeit ausgetauscht werden, und dass es unmöglich ist, sich einen Überblick durch manuelle Sichtung zu verschaffen. Darüber hinaus muss die Vertrauenswürdigkeit der Informationen im Zuge des Abgleichs verschiedener Informationsquellen und Modalitäten eingeschätzt werden. Für diese Aufgaben ist es dringend erforderlich, die Sicherheitsbehörden durch automatisierte und KI-basierte Systeme bei der Informationssichtung und Überblickserstellung zu unterstützen.

Die Herausforderung für Sicherheitsbehörden liegt dabei in der Bewältigung der Informationsflut, welche durch die stetig anwachsende Datenlast einhergeht. Personell können Behörden dem nicht mehr entgegenwirken. Deshalb ist ein übergeordnetes Ziel von HYBRIS, unstrukturierte Daten mit Hilfe Künstlicher Intelligenz so zu strukturieren, dass ein möglichst breiter Überblick über wesentliche Informationen gewährt werden kann und Zusammenhänge leicht erkennbar sind. Hierfür sollen Ansätze zur Erkennung relevanter Narrative und deren Bedrohungspotential für kritische Infrastrukturen erforscht werden. Die Erkennung von Narrativen und deren Kontext, ermöglicht es Sicherheitskräften einen nötigen Handlungsspielraum zu erlangen, um entsprechend auf Desinformation und daraus resultierenden Gefahren reagieren zu können.

HYBRIS wird in enger Zusammenarbeit mit den österreicheschen Sicherheitsbehörden erforscht. Erkenntnisse, Erfahrungen und entworfene Konzepte werden in einem Entscheidungsgrundlagen Dokument den jeweiligen Sektionen und Behörden zur Verfügung gestellt. HYBRIS soll prototypisch für eine nationale Data Intelligence Plattform erforscht werden, anhand derer Empfehlungen zur Verbesserung der Resilienz der österreichischen Sicherheitsbehörden hinsichtlich Hybrider Bedrohungen und konzertierter Desinformation bereitgestellt werden.

Abstract

Although disinformation has been a recognized problem for some time now and is actively dealt with at national and international level, it continues to grow in relevance and effect on a daily basis - from disinformation-led voter manipulation, to willful influence on the pandemic response, to targeted war propaganda. The latter has also drawn attention to a new phenomenon: hybrid threats. This abstract aggregation of different threats thwarts traditional responsibilities of Austrian security agencies, in that they are by definition carried out by state or non-state actors through a combination of overt and covert military and non-military means. As a result, these fall under the purview of either the BMI (e.g., cyber security, cyber crime), the BKA (e.g., disinformation, public safety), or the BMLV (e.g., cyber defense). A concerted hybrid threats attack therefore requires the cooperation of security agencies, and the facilitation of cooperation and the exchange of analysis results and counter strategies is the starting point of HYBRIS.
HYBRIS addresses hybrid threats that have the intent to influence beliefs and attitudes of selected audiences through online coordinated operations, mobilize them to act, and subsequently compromise physical and digital infrastructure. Disinformation is an essential component of hybrid threats and a specific focus in the project. Disinformation campaigns are often aimed at stoking fears in the population, which can result in various consequences. Whether or not messages are disinformation, it is important for security agencies to identify whether reactions to messages on social media and other news channels, or potentially subsequently organized actions, pose a threat to people or critical infrastructure. Only when it comes to taking appropriate action, in the case of the spread of disinformation, is it also a matter of countering a

threatening development by creating and disseminating counterstatements based on facts and trustworthy information. Responding quickly to this type of news trend is particularly important. Therefore, preparation of the content of possible counterstatements must begin while disinformation is still in the making. One of the main challenges in this case is to ensure a comprehensive overview of the currently prevailing trends. The difficulty is that information is exchanged at high speed through an unmanageable number of information channels, and it is impossible to obtain an overview by manual verification alone. In addition, the trustworthiness of information must be assessed in the course of matching different information sources and modalities. For these tasks, there is an urgent need to support security agencies with automated and AI-based systems for information prioritization and overview generation.
The challenge for security authorities is to cope with the flood of information resulting from the ever-increasing data load. In terms of personnel, authorities can no longer counteract this. Therefore, a primary goal of HYBRIS is to structure unstructured data with the help of artificial intelligence in such a way that the broadest possible overview of essential information can be provided and correlations are easily recognizable. For this purpose, approaches for the detection of relevant narratives and their threat potential for critical infrastructures will be explored. The recognition of narratives and their context enables security forces to gain the necessary room for maneuver in order to react appropriately to disinformation and the resulting threats.
HYBRIS is being researched in close cooperation with the Austrian security authorities. Results, experiences, and designed concepts will be made available to the respective sections and authorities in a decision support document. HYBRIS will be researched as a prototype for a national data intelligence platform, which will be used to provide recommendations for improving the resilience of the Austrian security authorities with regard to hybrid threats and concerted disinformation.

Endberichtkurzfassung

Das Projekt HYBRIS zielte darauf ab, bedrohliche Narrative und deren potenzielle Auswirkungen auf kritische Infrastrukturen zu erkennen, Desinformation und hybride Bedrohungen zu analysieren sowie die Reaktionsfähigkeit und Zusammenarbeit von Sicherheitsbehörden zu verbessern. Durch den Einsatz moderner Machine-Learning-Modelle und automatisierter Analyseverfahren konnten Methoden zur Erkennung relevanter Narrative erstellt und Ansätze zur Bedrohungslage gestet werden. Ein zentrales Ergebnis war die Entwicklung von Modellen zur frühzeitigen Erkennung von Desinformationskampagnen, darunter Themen wie Hassrede, Toxizität und Berichtsstilen.

Zur effizienten Analyse großer Datenmengen wurden automatisierte Werkzeuge zur Informationssuche, Klassifikation und Visualisierung von Zusammenhängen entwickelt. Eine zentrale Komponente war die Erstellung einer Taxonomie für kritische Infrastrukturen, mit der Inhalte gezielt analysiert und kategorisiert werden können. Durch den Einsatz von Knowledge-Graphen und semantischer Suche konnten Beziehungen zwischen Narrativen und Themen besser nachvollzogen werden. Zusätzlich wurde ein Prototyp für eine Data Intelligence Plattform erstellt, um die interaktive Exploration von großen Datenmengen zu ermöglichen.

Ein wesentlicher Bestandteil des Projekts war die rechtliche und ethische Bewertung der entwickelten Methoden. Datenschutzrechtliche Rahmenbedingungen wurden umfassend analysiert, insbesondere in Bezug auf nationale und EU-weite Regularien. Zudem wurde ein Ethics Impact Assessment Tool entwickelt, das ethische Risiken wie Diskriminierung oder algorithmische Verzerrungen identifiziert und bewertet. Die Forschungsergebnisse wurden aktiv mit nationalen und internationalen Bedarfsträgern abgestimmt, um eine bestmögliche Integration in bestehende Sicherheits- und Analyseprozesse zu gewährleisten.

Neben der technischen und wissenschaftlichen Forschung wurde besonderer Wert auf die Verbesserung der Zusammenarbeit zwischen Sicherheitsbehörden gelegt. Durch regelmäßige Abstimmungen und Workshops mit relevanten Institutionen konnten die entwickelten Werkzeuge praxisnah getestet und optimiert werden. Die erzielten Ergebnisse tragen dazu bei, die Effektivität und Resilienz von Sicherheitsbehörden zu steigern und eine fundierte Grundlage für zukünftige Entwicklungen im Bereich der Bedrohungsanalyse und Desinformationsbekämpfung zu schaffen.

Das Projekt wurde durch eine gezielte Öffentlichkeitsarbeit begleitet, um die wissenschaftlichen Erkenntnisse einem breiten Publikum zugänglich zu machen. Dies umfasste Publikationen, Fachkonferenzen und öffentliche Veranstaltungen, bei denen die Forschungsergebnisse präsentiert und diskutiert wurden. Insgesamt konnten alle definierten Projektziele erreicht werden, wodurch ein wichtiger Beitrag zur Stärkung der digitalen Sicherheit und der Bekämpfung hybrider Bedrohungen geleistet wurde.

Projektkoordinator

- AIT Austrian Institute of Technology GmbH

Projektpartner

- Technische Universität Wien
- Artificial Researcher IT GmbH
- thinkers GmbH
- Bundesministerium für Landesverteidigung
- Universität für Bodenkultur Wien
- Research Institute AG & Co KG