AUSTRIAN INSTITUTE OF TECHNOLOGY

RESEARCH FOCUS DISINFORMATION DETECTION

Cross-project, intersectoral linkages and coordination

Version 02, 28.06.2023

DISINFORMATION DETECTION

PROJECT LINE

AUSTRIAN INSTITUTE OF TECHNOLOGY

Study on Desinformation Detection

- **Overview of technological options** to counter desinformation
- First Tech-Pilot

RAIDAR — RAPID AI BASED DETECTION OF AGGRESSIVE OR RADICAL CONTENT ON THE WEB

- Analysis of social media channels with regard to **Hate Speech** and **Extremist content**
- Approaches to fight **Infodemic** (support in coping with information overload)
- **Hate Speech** and **Toxic Content** Analysis (e.g., Sexism, toxicity, discrimination)
- **Extremist Content** Analysis (e.g., political, religious, criminal relevance)

STARLIGHT

- **Easy deployable Tools for LEAs**
- Image manipulation Detection
- Text Content Analysis

2019 — 2020 — 2021 — 2022 — 2023

defalsif-ai

- Developed a large **Medi-Forensics Toolbox**
- Audio-Visual forensics to facilitate Fact Checking
- **Audio Tampering** Detection
- **Image/Video manipulation** Detection
- **Deep Fake** Detection
- **Text content analysis** (e.g., writing/reporting style, act claiming, propaganda)

German-Austriar Digital Media Observatory — GADMO

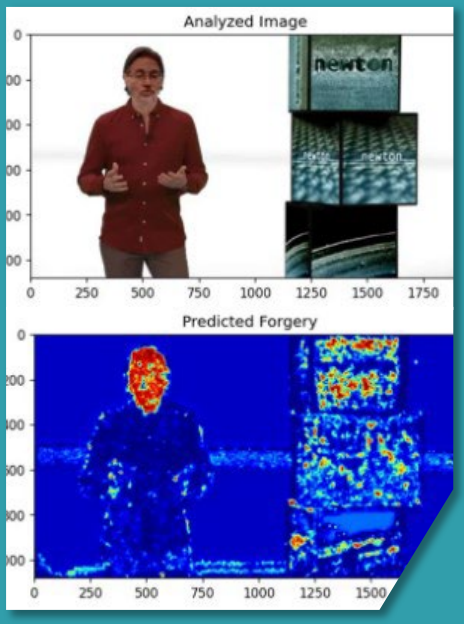
- Detecting and analysing disinformation campaigns
- support mainstream, local media and public authorities in exposing harmful disinformation campaigns
- Organizing media literacy activities at ´ national or multinational level
- Providing support to national authorities for the monitoring of online platforms' policies and the digital media ecosystem

HYBRIS

- Identification and Analysis of **Hybrid Threats**
- **Large Scale** Desinformation **Trend Analysis**
- **High Performance Machine Learning** Stacks
- Detection of **Narratives**
- Improved **Infodemic** support

TARGET SETTING

- Detection of manipulation in media
- Detection of artificially created media and deepfakes
- Methods for traceability and provability when using AI methods to detect fake news
- Analysis of the legal situation and the possibilities to take action against e.g. deepfakes.



Detecting Deep Fake Manipulation in Videos

PROJECT LINE DISINFORMATION DETECTION

AUSTRIAN INSTITUTE OF TECHNOLOGY

STUDY	DEFALSIF-AI	RADAR	HYBRIS	FAKE NEWS EARLY WARNING SYSTEM
Tasks and threat area				
Study on threat technologies, Counter-measures, investment strategy,Recommendation catalog	Detection of disinformation, audiovisual media manipulation, text content analysis	Detection of hate on the network, radicalization, democracy-threatening content, threat potential analysis	Detection of disinfor-mation campaigns in Big Data streams. Resilience to Hybrid Threats	Multi-stake-holder platform: "Weather service" for fake news trends. Knowledge base on disinformation
Application areas				
• Individual files	• Individual files • Web-URLs	• Individual social media Channels • Confiscated hard drive, Cell phones	• Variety of different Social media channels • Different heterogeneous sources	• Unlimited number of heterogeneous channels, sources and content
Analysis and detection				
• First Deep Fake Recognition prototype	• Manipulations in image and sound • Deep fakes • Extensive text analyses	• Hate Speech • Text Analysis: • Sexism, antisemitism, radicalism. • Radical symbolism	• Fake News Narrative • Topic detection / Trend analysis • Automatic Summary	• Trans-national / Cross-source Trend analysis • Cluster analysis
Understanding / Knowledge acquisition / Trend identification				
• Overview of threat situations and technical possibilities	• Recognizing and explaining image and audio manipulations	• Gaining overview of topics and content in larger channels	• Fake News Narrative (Monolingual) • Local Fake News Trends	• Multilingual Narrative Fusion • Global Fake News Trends
Results				
• Reports • Recommendation catalog	• Analysis platform for media forensics	• Analysis platform for data streams	• Big Data / HPC analysis platform	• Online platform for fake news trends

AI-based fact-checking tools

AUSTRIAN INSTITUTE OF TECHNOLOGY

Approach

- Provide tools to support fact-checkers
- Media forensic detection of manipulation
- Recognition of synthetic content

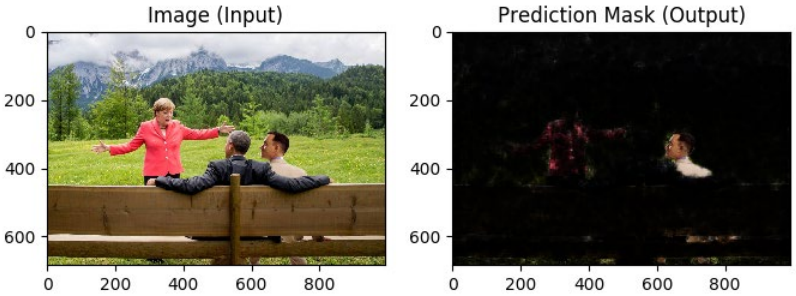


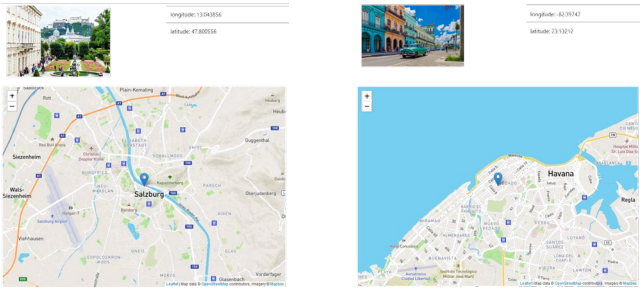
Image manipulation detection

AI-based recognition of whether something has been manipulated - inserted / deleted - in an image. Clear presentation of the analysis results. The image on the right shows what has been added to the image on the left.

Recognising the recording location

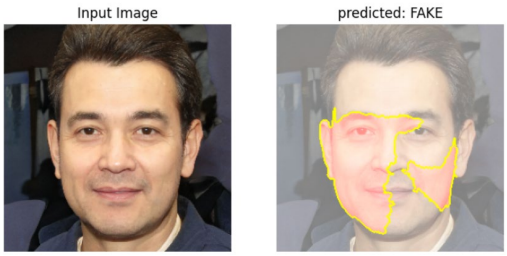
It is often important to check whether a picture was actually taken at the specified location.

For this purpose, models have been developed that can determine the location of the recording. This method works very well at known locations, but also in open terrain with an accuracy of up to 100 km deviation.



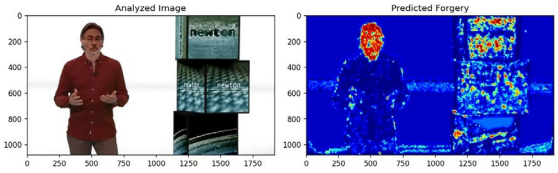
Recognise fake profile photos

Fake profiles in social media are becoming an increasing problem. Generative models can be used to create better and better fake profile images. Our neural network was trained with 125,000 images from various sources and achieves a correctness of 95-99.8 % on benchmark data sets.



Detecting Deep Fakes

Fake profiles in social media are becoming an increasing problem. Generative models can be used to create better and better fake profile images. Our neural network was trained with 125,000 images from various sources and achieves a correctness of 95-99.8 % on benchmark data sets.



TEXT CONTENT ANALYSIS

AUSTRIAN INSTITUTE OF TECHNOLOGY

Challenge

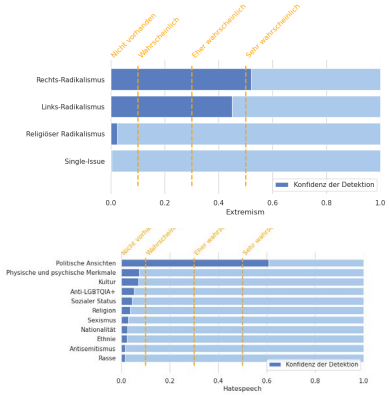
- Direct recognition of disinformation often hardly possible
- Requires broad general knowledge (not available in AI)

Approach

- Determination of several relevant content descriptions and characteristics
- Presentation by means of **Information Nutrition Labels**
- **Multi-modal fusion of** the features into an overall assessment with regard to the (dis-) information content.

AI MODELS for content description

- Each content feature is derived from the online data by a separate AI module.
- Description of the (des-) information content.
- Portfolio of AI modules developed over several projects (see table)

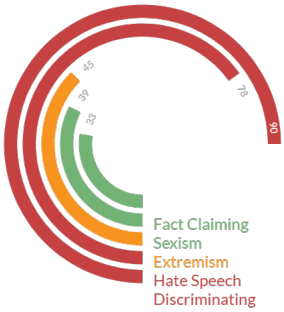


Comprehensible presentation

A clear and concise presentation of results and information is also the focus of research activities. New approaches to visualisation are being researched for this purpose.

Information Nutrition Labels

describe the content of documents or online articles in a clear way. Users get a quick assessment of the information content.



Explainability of AI

Explainability and simple comprehensibility are central requirements for AI modules. The user must always be able to interpret the AI's decisions and assessments.

Text with highlighted words

Ein typischer Wirtschaftsflüchtling. Ab nachhause mit ihm.Abgesehen davon: Niemand hat ein Problem mit solchen Menschen, solange der Staat für seine Bürger, also für jene, die dafür auch bezahlen, gut funktioniert. Das tut er aber nicht.Kriegen unverschuldet obdachlose Österreicher auch ein Zelt?

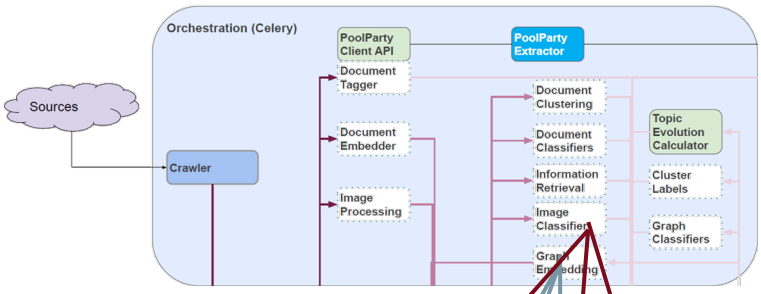
Name	Recognised contents	Language	Domain	Category Examples
Fake News	Direct detection of fake news	English	Social networks	Yes / No
Hate speech	Hatred against groups or individuals	Multi-ling	Social networks Discussion forums	Yes / No
Extremism	Extremist content	German	Social networks Article	Right-, Left-, Religious- or Single-Issue Extremism
Toxicity	Toxic, offensive content, comments, hateful language	German	Social networks	Yes / No
Factual assertions	Was it factually alleged?	Multi-ling	Social networks	Yes / No
Appealing contents	Appealing, positive, discussion-promoting, language	German	Social networks Article	Yes / No
Sentimentality	Sentiment, feeling, emotion	German	Article	Positive, Negative
Report style	Report style of an article	German	Article	Conspiracy theory, clickbait
Writing style	Writing style of an article	German	Article	Polarise, exaggerate
Discrimination	Is a statement discriminatory?	German	Social networks	Ethnicity, social status
Relevance to criminal law	Is a statement criminal?	German	Social networks	Incitement, insult
Sexism	Various categories of sexism	English	Social networks	Misogyny, Sexual Violence

FAKE NEWS TREND ANALYSIS

AUSTRIAN INSTITUTE OF TECHNOLOGY

Privacy Aware Data Acquisition

Intelligent crawlers for different social networks and platforms, which automatically obtain relevant data while taking data protection into account.

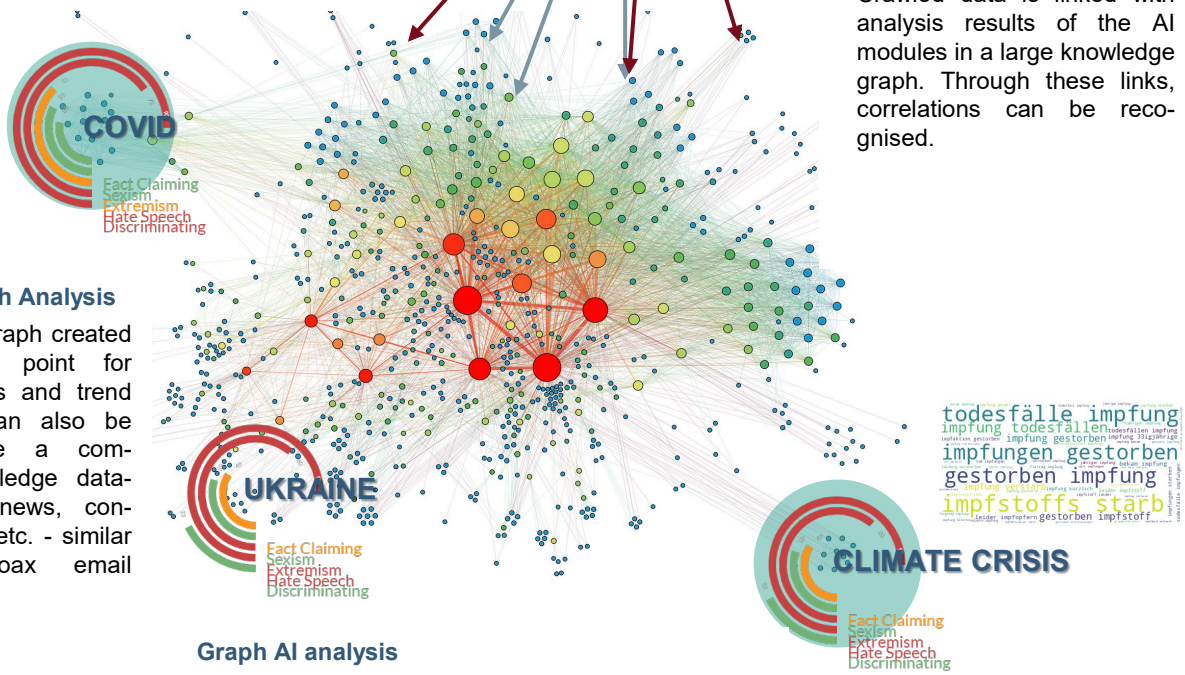


Complex AI pipelines

Disinformation is complex and requires many specific AI modules for detection. Each item is analysed by a multitude of modules. The efficient management of such complex pipelines requires optimal planning and ingenuity.

Information networking

Crawled data is linked with analysis results of the AI modules in a large knowledge graph. Through these links, correlations can be recognised.



Knowledge Graph Analysis

The knowledge graph created is the starting point for complex analyses and trend predictions. It can also be used to create a comprehensive knowledge database on fake news, conspiracy theories, etc. - similar to existing hoax email databases.

Graph AI analysis

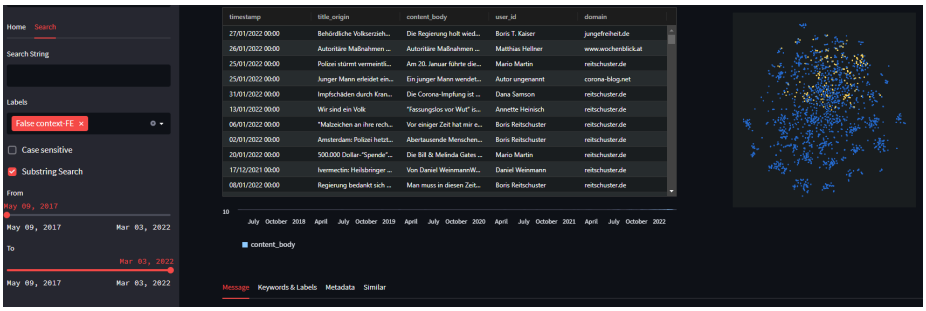
Graph Neural Networks are the latest trend in the field of artificial intelligence. This promising technology makes it possible to model and evaluate highly complex correlations. Especially for such complex and subjective tasks as the interpretation of (dis-) information content, they represent an optimal solution to link the different data formats (text, image/video, sound, relationships in social networks, etc.) with each other, or to automatically recognise links.

Network analysis

Detection of distribution channels and key actors in disinformation networks. Detection and analysis of echo chambers and bot networks.

Clear presentation of topics

Topic clusters visualised by means of ***Information Nutrition Labels.*** Quick overview through automatically extracted ***keywords*** and ***short summaries.***



Data Exploration Tool - Result Project RAIDAR (FFG KIRAS)

Trend analysis in global news - result project STARLIGHT (EU H2020)

INFODEMIC COMBAT

AUSTRIAN INSTITUTE OF TECHNOLOGY

Too much information through too many channels

Infodemic describes the powerlessness in the face of the permanent flood of news, in which it is no longer possible to distinguish whether something is true or false.

Approach

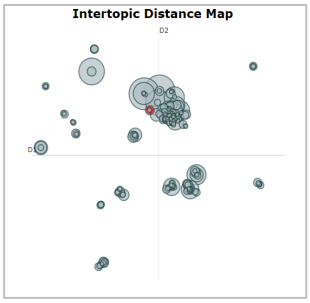
- Structure content automatically
- Summarise relevant content from large amounts of news
- Clear information visualisation
- Show relationships and similarities

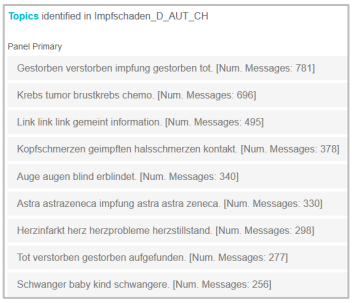
Infodemic is

"… an overabundance of information – some accurate and some not – that makes it hard for people to find trustworthy sources and reliable guidance when they need it"

Themes Detection

Automatic recognition of connections based on text similarity and semantic analysis. Clear presentation of topic clusters and their similarities. Hirarchical structure in sub-topics.



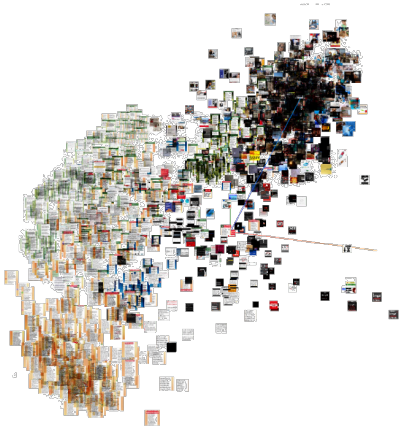


Keyword recognition

Automatic recognition of relevant keywords. Enable a quick overview of the content of an article or one or more social media channels.

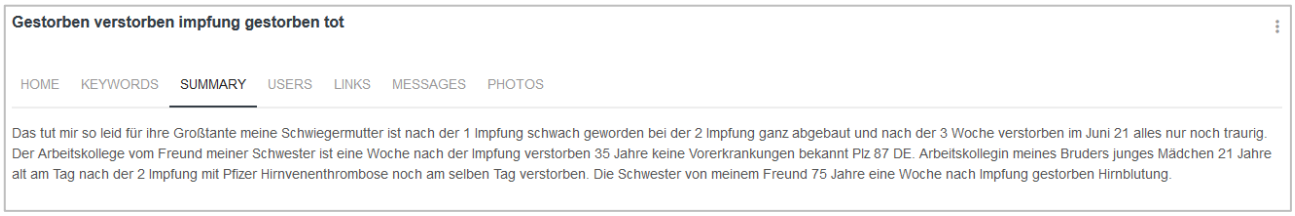
Representation of semantic similarity

Calculate and display similarities in media collections - e.g. images, texts, videos - so that users can better recognise connections.



Automatic short summary

Short summary of one or more articles to get a quick overview of shared content or discussions.



COOPERATION PARTNER

MINISTERIAL COOPERATION

- Federal Chancellery
- Federal Ministry Republic of Austria Justice
- Federal Ministry Republic of Austria European and International Affairs
- Federal Ministry Republic of Austria Defence
- Federal Ministry Republic of Austria Interior

INSTITUTIONAL COOPERATION

ORF
APA AUSTRIAPRESSEAGENTUR
dpa Deutsche Presse-Agentur GmbH
AFP
CORRECTIV Recherchen für die Gesellschaft

RESEARCH AND INDUSTRY PARTNERSHIPS

LIquA Linzer Institut für qualitative Analysen
research institute
SCENOR THE SCIENCE CREW
Donau-Universität Krems Universität für Weiterbildung
thinkers.ai Best in results
TU WIEN TECHNISCHE UNIVERSITÄT WIEN Vienna University of Technology
BOKU Universität für Bodenkultur Wien
enliteAI
SEMANTIC WEB COMPANY
Artificial Researcher
VIENNA SCIENTIFIC CLUSTER
ATC ATHENS TECHNOLOGY CENTER
tu technische universität dortmund

FUNDING PROGRAMS

FFG Promoting Innovation.
KIRAS Security Research
European Commission | Horizon 2020 European Union funding for Research & Innovation

Contact

ALEXANDER SCHINDLER
Thematic Coordinator Datascience
Data Science & Artificial Intelligence
Center for Digital Safety & Security

AIT Austrian Institute of Technology
Giefinggasse 4 | 1210 Vienna | Austria
+43 664 8251454
alexander.schindler@ait.ac.at

MARTIN BOYER
Senior Research Engineer
Sensing & Vision Solutions
Center for Digital Safety & Security

AIT Austrian Institute of Technology
Giefinggasse 4 | 1210 Vienna | Austria
+43 664 8251440
martin.boyer@ait.ac.at

MICHAEL MÜRLING
Marketing and Communications
Center for Digital Safety & Security

AIT Austrian Institute of Technology
Giefinggasse 4 | 1210 Vienna | Austria
T +43 50550-4126
michael.muerling@ait.ac.at