Press release

Vienna, 19.03.2024

FIGHT AGAINST DEEPFAKES: NEW PROJECT TO DETECT IMAGE AND VIDEO MANIPULATION

New KIRAS project - led by the AIT Austrian Institute of Technology - focuses on improving the detection and combating of deepfakes in digital image and video content

Whether it's DALL-E 3, face filters on TikTok and Instagram or DeepFaceLab: technologies that enable the manipulation of videos and images are booming, are increasingly easy to access and mark a turning point for many people when it comes to trusting digital content. How to recognize what is real when the fakes are increasingly realistic? From public authorities and administrations to media organizations, the private sector and civil society, everyone is confronted with the challenges and dangers of deepfakes.

Focus on tool development and preventive work

The "Defame Fakes" project, coordinated by specialists in artificial intelligence (AI) and image and video analysis at AIT and funded by the Federal Ministry of Finance's KIRAS security research funding program, focuses on researching and developing suitable and effective tools to support the semi-automated detection of deepfakes in large data sets. Preventive awareness measures are also intended to initiate a discourse in society as a whole and raise awareness of the problem. The aim is to counteract the continuous erosion of trust in digital content and, by creating new technological possibilities for detecting image and video manipulation, to protect companies and society against manipulation, create awareness in dealing with digital information and thereby strengthen trust in digital media. The project is funded by the Federal Ministry of the Interior (BMI) and the Federal Ministry of Defense (BMLV).

Disinformation in the context of hybrid threats

"Democracy and the democratic formation of opinion are currently being called into question to an unprecedented extent. AI allows misleading content to be created quickly and disseminated with a wide reach," explains Michael Suker, Head of the Cyber Documentation & Research Center/ZentDok at the National Defense Academy. There is a risk that deepfakes could be received unnoticed by the population in order to influence public opinion. Deepfakes can be part of hybrid threats aimed at undermining or damaging the political, economic and social structure of a state or organization by combining various means. These include cyberattacks, disinformation and economic pressure. Suker continues: "Detecting and combating deepfakes and improving media literacy are therefore crucial to protect the integrity of information and ensure national and international security."

Analysis of social, ethical and legal consequences

The analysis of the social, ethical and legal implications of the spread of deepfakes and the use of detection tools plays a central role in all research activities: What does it mean for corporate security if the voice and image of a CEO can be imitated using deepfake technology? How can private individuals be protected preventively against fraud, which is made even more credible by generative AI? What consequences do disinformation, the manipulation of opinion-forming processes or the promotion of extremist groups through deepfakes have for democracy? And what regulation is needed not only to meet the challenges of deepfakes, but also to make the use of detection tools safe?

Interdisciplinary team researches the detection of deep fakes

The project is coordinated by AIT, which is now an established competence center in the field of AI-based applications for protection against disinformation campaigns - both nationally and internationally. The innovative solutions developed at AIT aim to protect media consumers and thus democracy as the cornerstone of our society. Martin Boyer, project manager and senior researcher at AIT: "As coordinator, it was important for me to put together a broad-based consortium in order to cover all the necessary expertise within the project. After all, when it comes to deepfakes, we are dealing with a wide range of challenges that can only be mastered on an interdisciplinary basis".

An interdisciplinary team is therefore behind defame Fakes. "In order to understand the influence of deepfakes on social contexts, it is crucial not to lose sight of the civilian population: Prevention activities and awareness measures must reach everyone - from young people to the elderly," emphasizes Louise Beltzung from the Austrian Institute for Applied Telecommunications (ÖIAT). ÖIAT will investigate the implications of deep fakes in this area, in close cooperation with PwC Austria, which is focusing on the potential threat to companies. "Deepfakes pose a serious risk to companies by jeopardizing the credibility and security of business communications. They open up new avenues for fraud, extortion and reputational damage, which underlines the need for effective measures to detect and defend against them," emphasizes Roland Pucher, Head of PwC Cybersecurity & Innovation Lab.

The aim is not only to better understand the effects of deepfakes, but also to develop suitable countermeasures to protect society from the dangers of image and video manipulation and restore trust in digital media. The research results should also help in practice - for example in training journalists. "It is important that, following the defalsif-AI project, we once again find ourselves in a competent consortium with different experiences and interests. We are pleased to be able to contribute our knowledge from everyday editorial work and to be able to work together to counter current challenges," says Florian Schmidt, head of the fact check team at APA - Austria Press Agency.

"This is a significant step in the right direction," says Alexander Janda, Secretary General of KSÖ - Kompetenzzentrum Sicheres Österreich, dissemination partner in the project: "The topic of deepfake is no longer just a dream of the future, but has arrived directly in the middle of society.

Recognizing manipulated image and video material, as well as raising public awareness, is therefore of the utmost importance."

About "defame Fakes"

"defame Fakes" is funded by the KIRAS security research funding program of the Federal Ministry of Finance. In close cooperation with the BMI and BMLV, the following partners are involved in the project:

- AIT Austrian Institute of Technology
- APA Austria Press Agency
- PwC Austria
- KSÖ Competence Center Secure Austria
- ÖIAT Austrian Institute for Applied Telecommunications

Further information can be found on the project website at: www.defamefakes.at.

Query note:

Mag. (FH) Michael W. Mürling
Marketing and Communications
AIT Austrian Institute of Technology
Center for Digital Safety & Security
T +43 (0)664 235 17 47
michael.muerling@ait.ac.at I www.ait.ac.at

Mag. Michael H. Hlava
Head of Corporate and Marketing Communications
AIT Austrian Institute of Technology
T +43 (0)50550-4014
michael.hlava@ait.ac.at I www.ait.ac.at