MALICIOUS USES AND ABUSES OF ARTIFICIAL INTELLIGENCE (AI)

David Sancho, Maria Eira, Aglika Klayn

Nov 2021

TREND MICRO™ | research

United Nations Interregional Crime and Justice Research Institute

CENTRE FOR ARTIFICIAL INTELLIGENCE AND ROBOTICS

EUROPOL EC3 European Cybercrime Centre

Background

Aim

- To provide an in-depth assessment of the present and possible future malicious uses and abuses of AI and related technologies
- To enhance awareness and preparedness to respond to such threats

Scope

- Malicious uses of AI
- Malicious abuses of AI
- Deepfakes case study

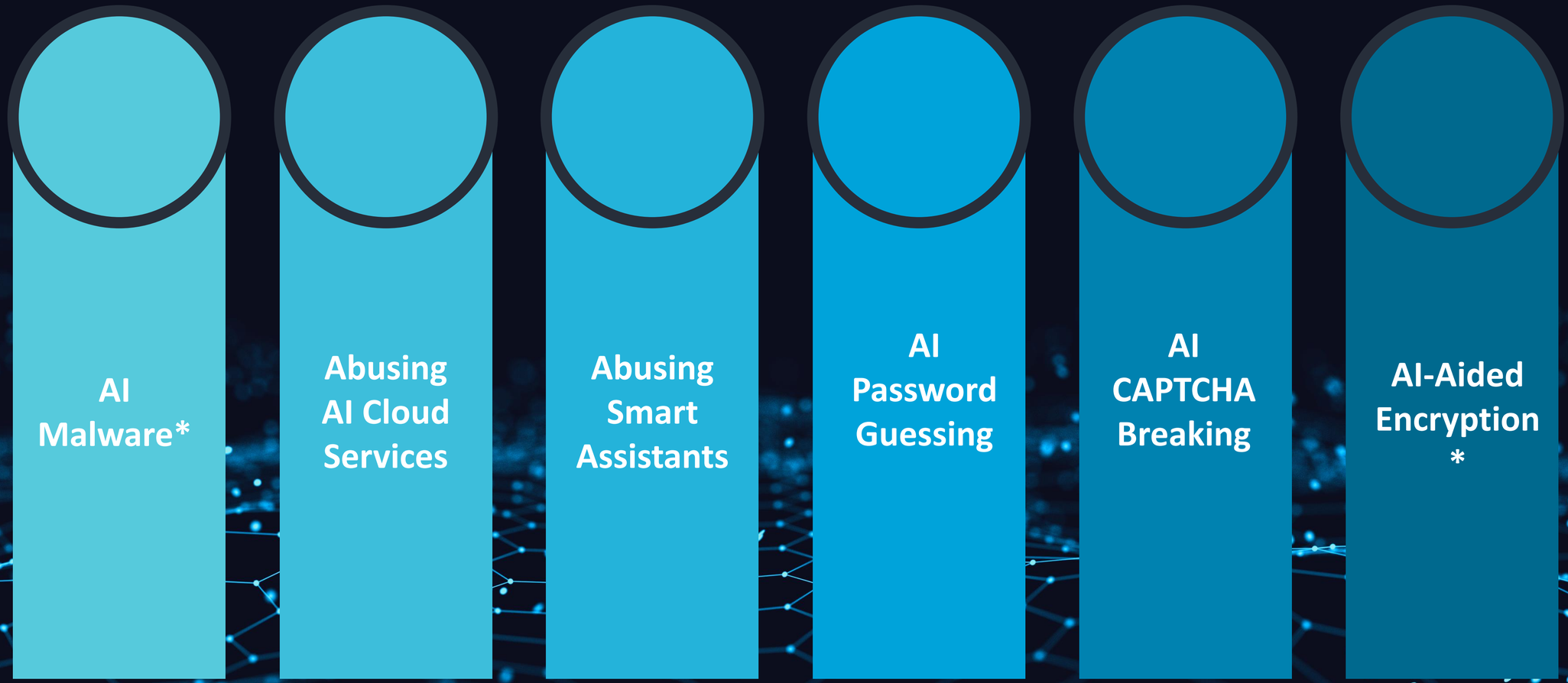
Methodology

- Research and contributions from the three entities
- Workshop held in March 2020

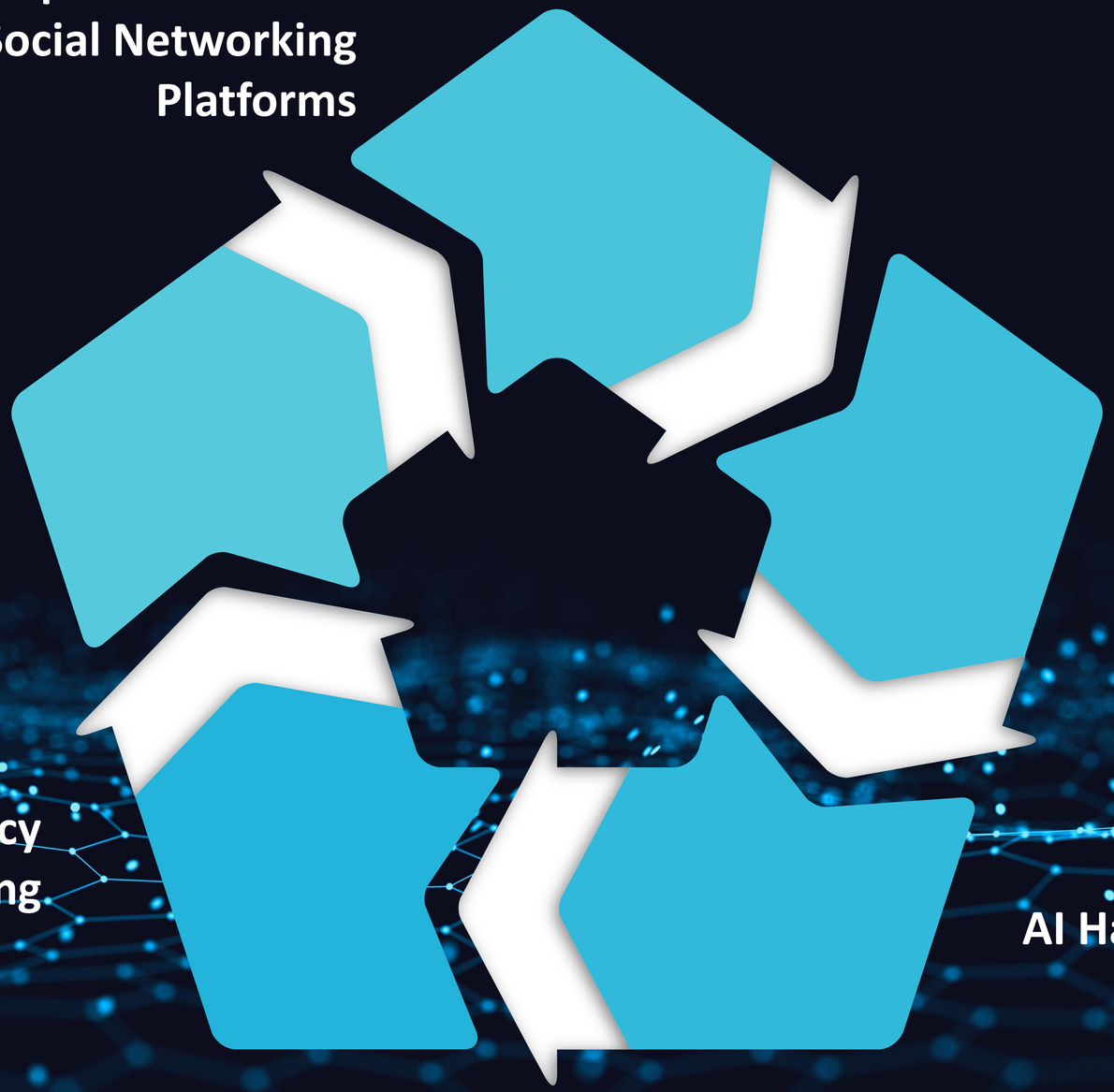
PART I

Present State of Malicious Uses and Abuses of AI

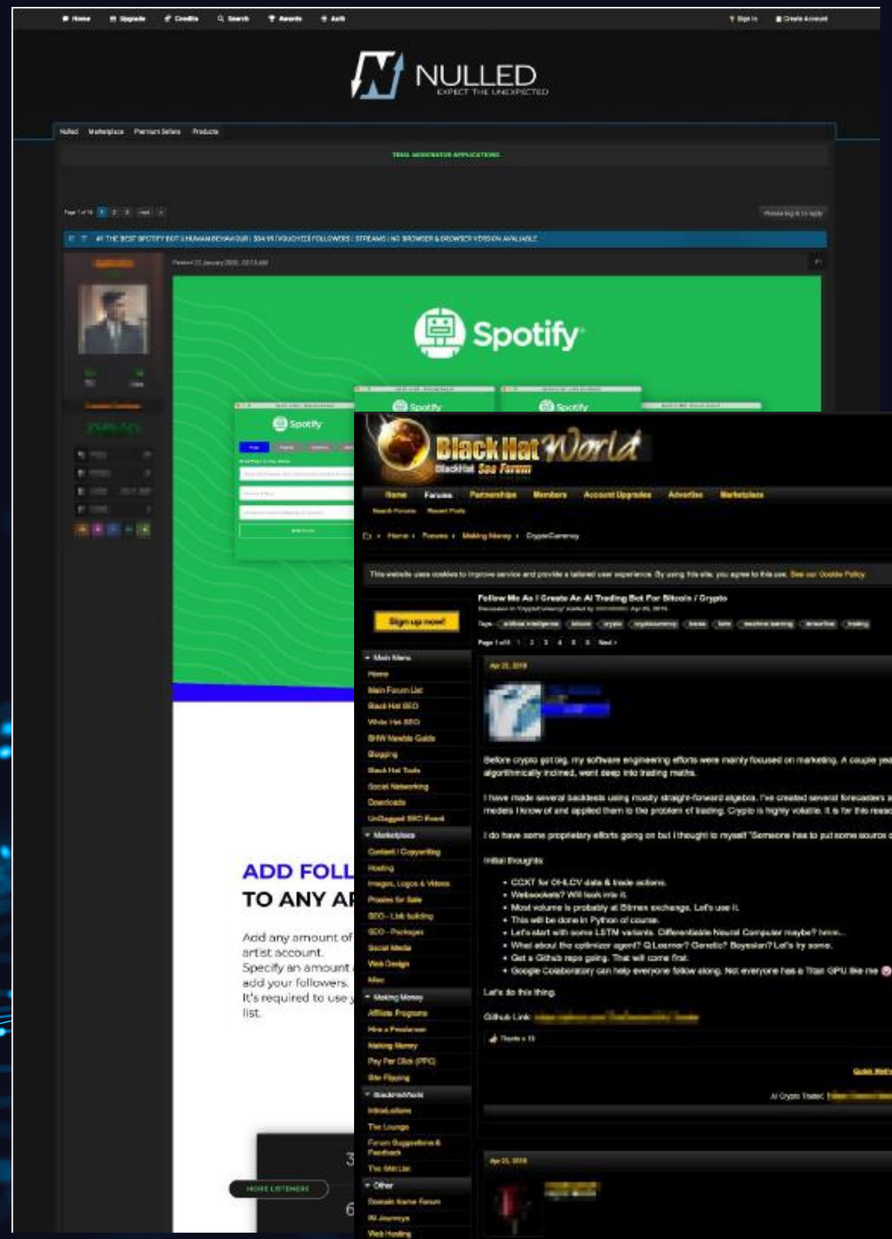
Present State of Malicious Uses and Abuses of AI

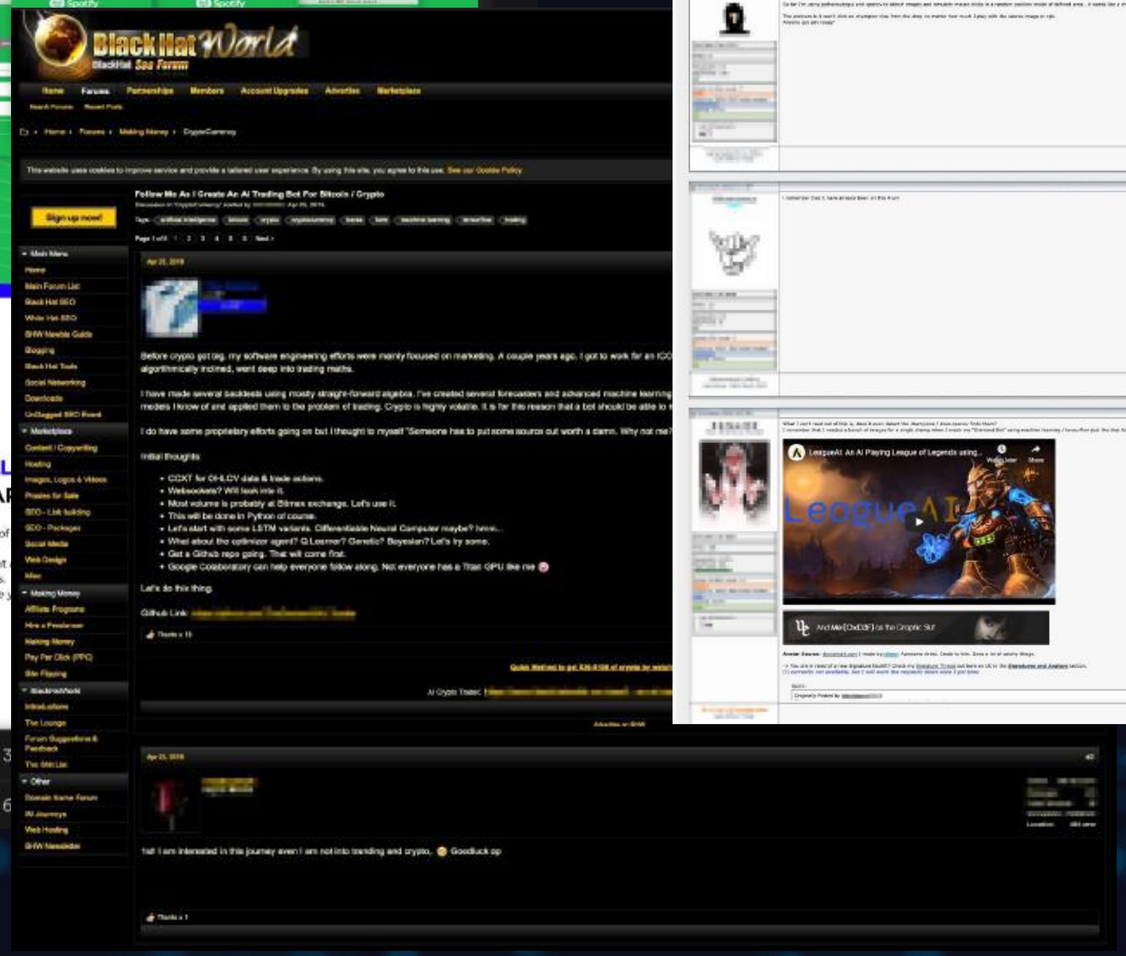


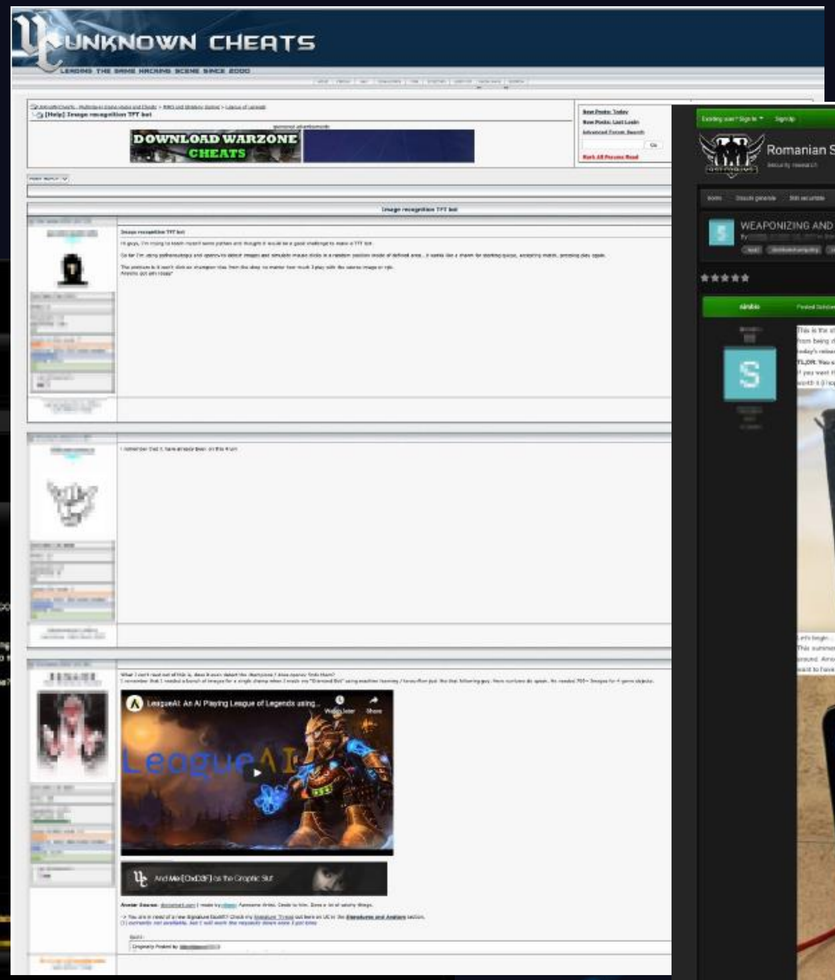
Trends Found on Underground Forums

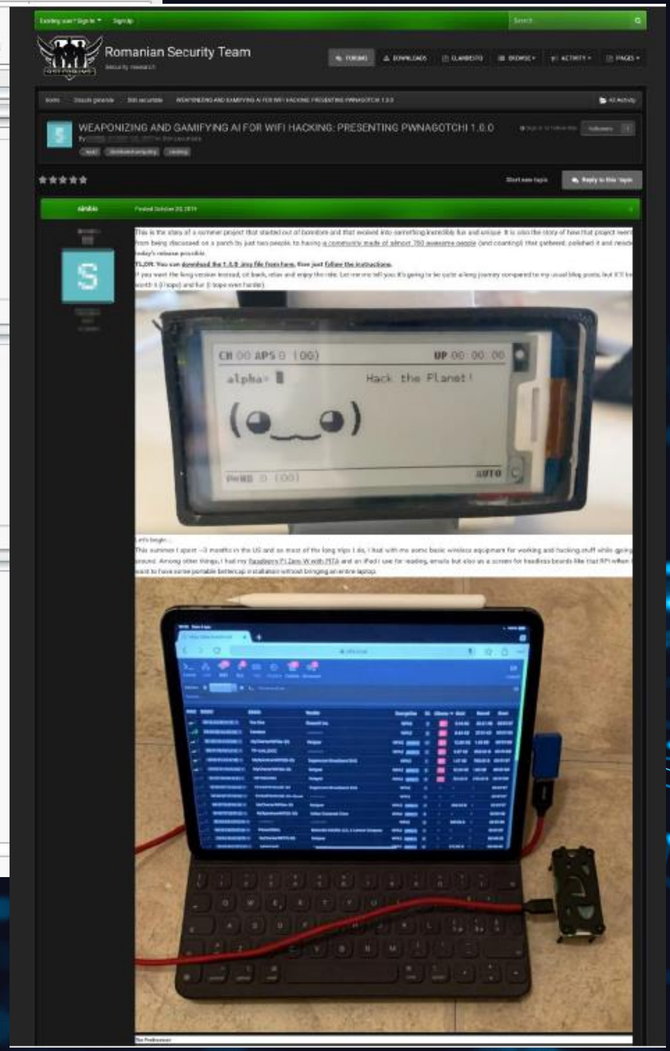


Examples from Underground Forums









PART II

Deepfakes

Deepfakes

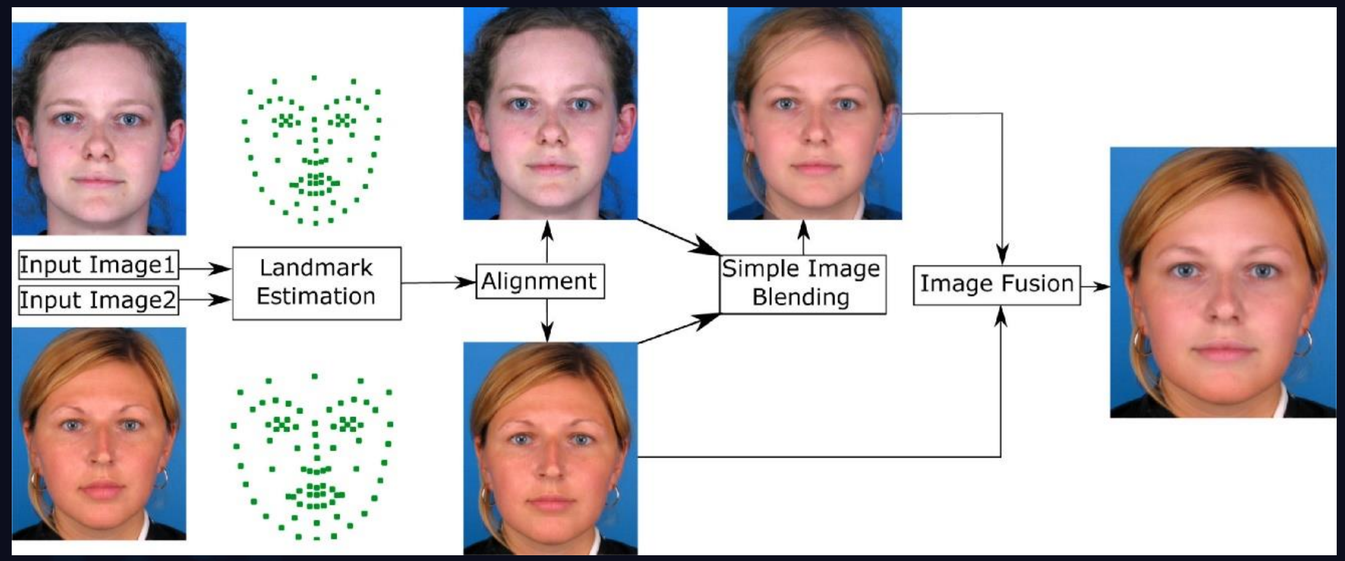
01 Different types of deepfakes & the technology behind them

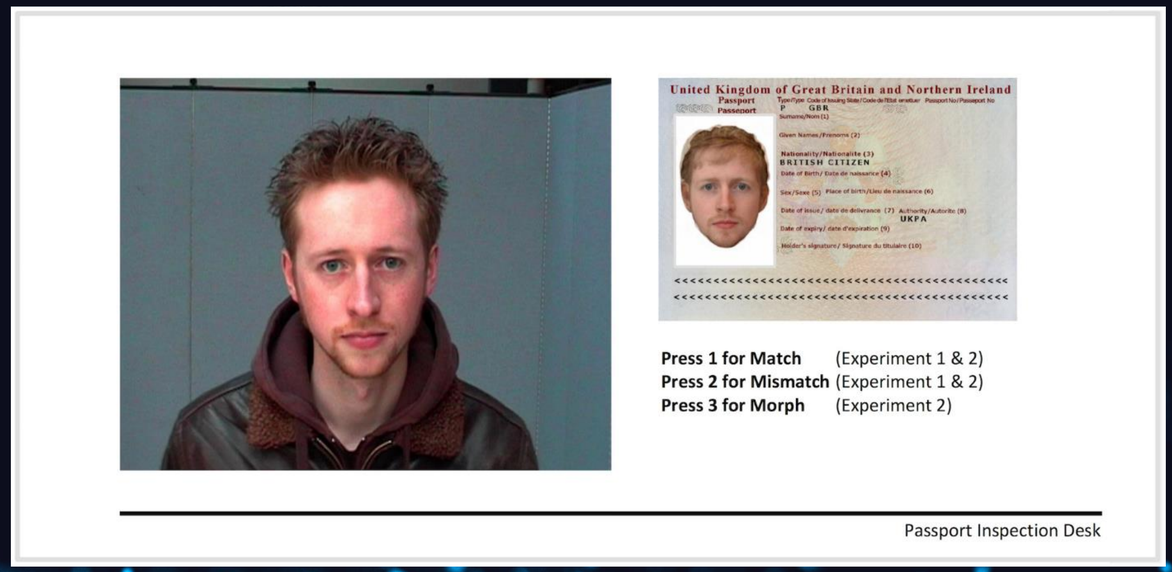
02 Current state of abuses of deepfakes & concrete examples

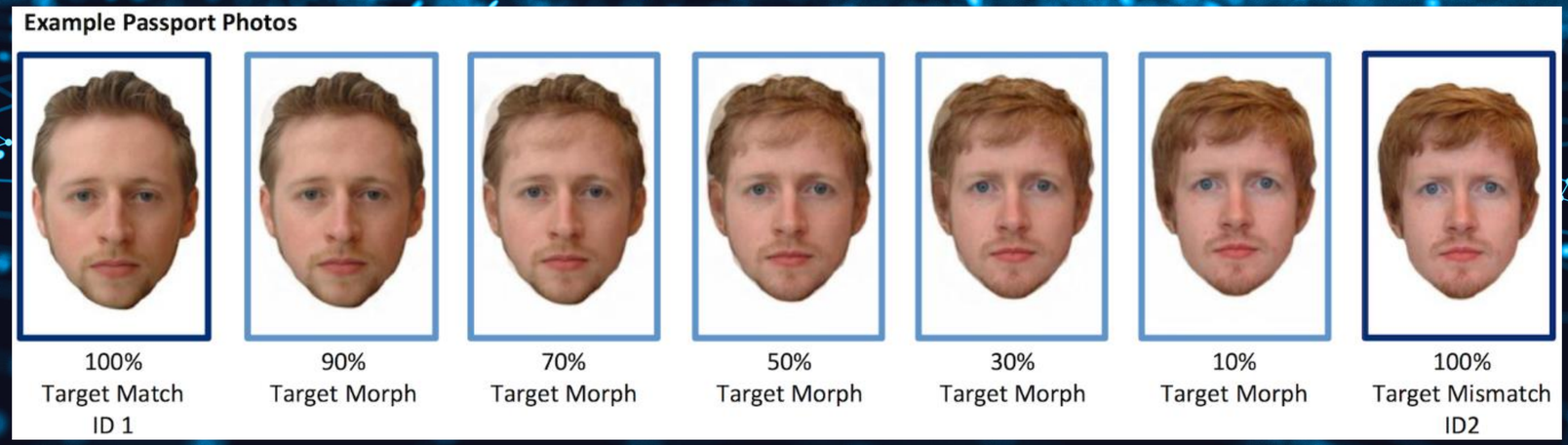
03 Possible future threats of deepfakes

04 Countering deepfakes approaches

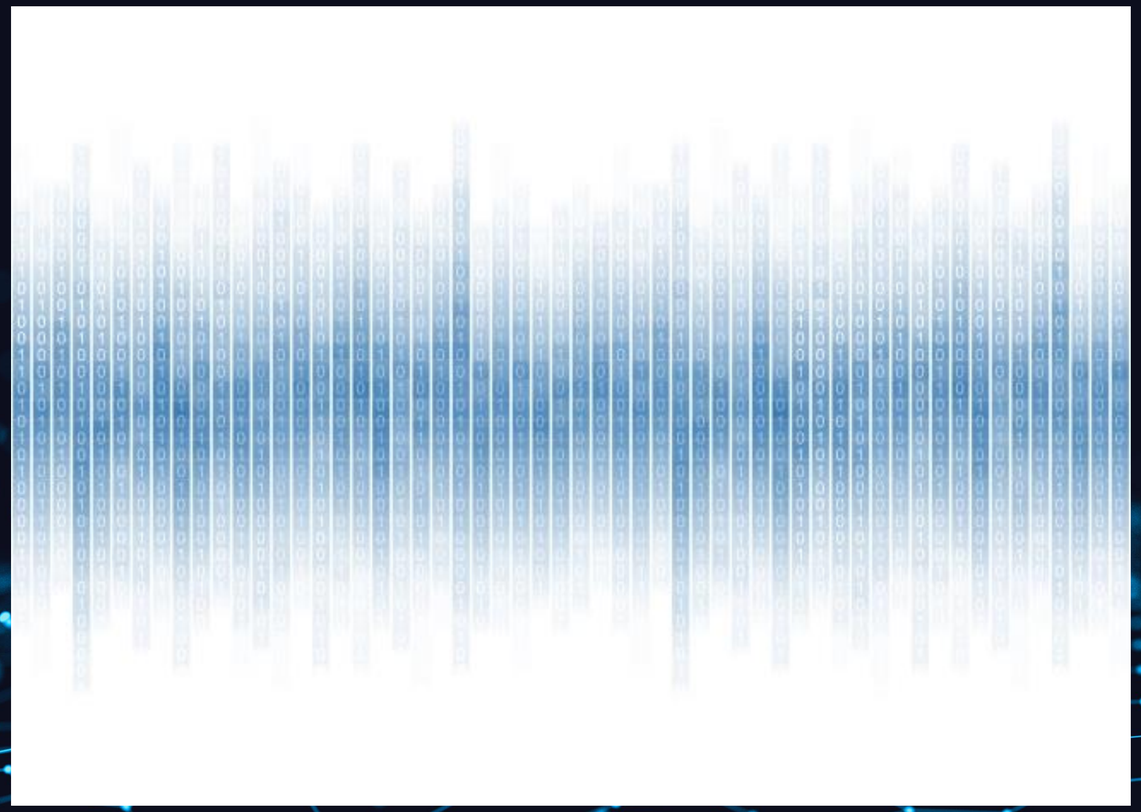
Face Morphing and Fraudulent IDs







Deepfake Voice Fraud Case Examples



AI-based voice fraud leads to $243,000 losses for energy company

AI voice cloning used in a $35 million bank heist

Possible Future Deepfake Threats

- Disinformation campaigns
- Securities fraud
- Extortion
- Online crimes against children
- Obstruction of justice
- Cryptojacking
- Illicit markets

PART III

Future Scenarios & Recommendations

Future Scenarios of Malicious Uses and Abuses of AI

Abusing Image Recognition Systems

Criminal Business Intelligence

Robocalling v2.0

Improved Social Profile Aging for Forums and Botnets

Content Generation and Parsing

Social Engineering at Scale*

Escaping an Image Recognition System

Remote Machine Learning Sets Pollution

AI-Enabled Stock Market Manipulation

Business Process Compromise

Insider Attacks (Banking & Trading Floor AI)

Local Library Poisoning by Resident Malware

AI-Supported Ransomware*

Escaping AI Detection Systems (Fraud & Voice Recognition in Banks)

Recommendations

AI for Good

- AI technology as a crime-fighting tool
- Responsible AI innovation
- Exchange of best practices

Further Research*

- Forward-looking threat assessments
- Continuous mapping of AI threat landscape
- Risk management driven threat classification & response

Secure AI Design Frameworks

- Security-by-design & privacy-by-design
- Specific data protection frameworks
- Human-centric approach to AI
- Technical standards

Policy

- Well-informed public debate
- International technology-agnostic policies
- Long-term sustainability driven by effective oversight

Outreach

- PPPs & multidisciplinary expert groups
- Enhanced AI literacy & cyber hygiene
- Knowledge sharing on tools, tactics, & techniques

TREND MICRO™ | research

United Nations Interregional Crime and Justice Research Institute

CENTRE FOR ARTIFICIAL INTELLIGENCE AND ROBOTICS

EUROPOL EC3 | European Cybercrime Centre

Thank you for the attention

Any questions?