sAIfer Lab
Joint lab on Safety and Security of AI

Improving Cybercrime Detection and Digital Forensics Investigations with Artificial Intelligence

Silvia Lucia Sanna –
University of Cagliari
National PhD in AI for Securtiy and Cybersecurity

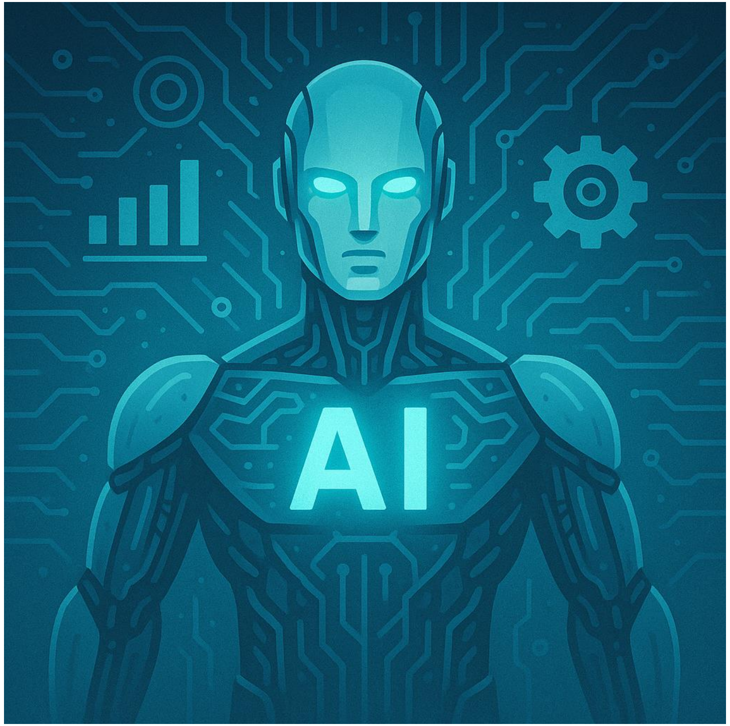
Cagliari, APWG Tech Summit June 2025

UNICA UNIVERSITÀ DEGLI STUDI DI CAGLIARI

Università di Genova

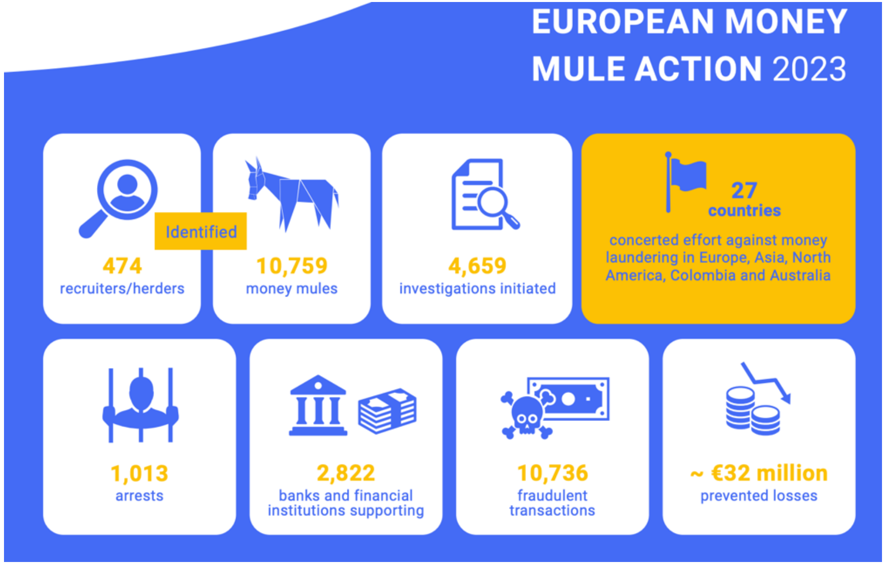
Overview

- Cybercrime and Digital Forensics
- Current Techniques
- AI for cybercrime detection
- AI for digital forensics investigation
- Case Study: general purpose LLMs for steganography



Introduction

- EUROPOL: Millions of European citizens are **daily victims of cybercrimes**
- EPPO: cybercrime **increase of 38% at the end of 2024** with damage of **24.8 billion euros**



EUROPEAN MONEY MULE ACTION 2023

Identified

474 recruiters/herders

10,759 money mules

4,659 investigations initiated

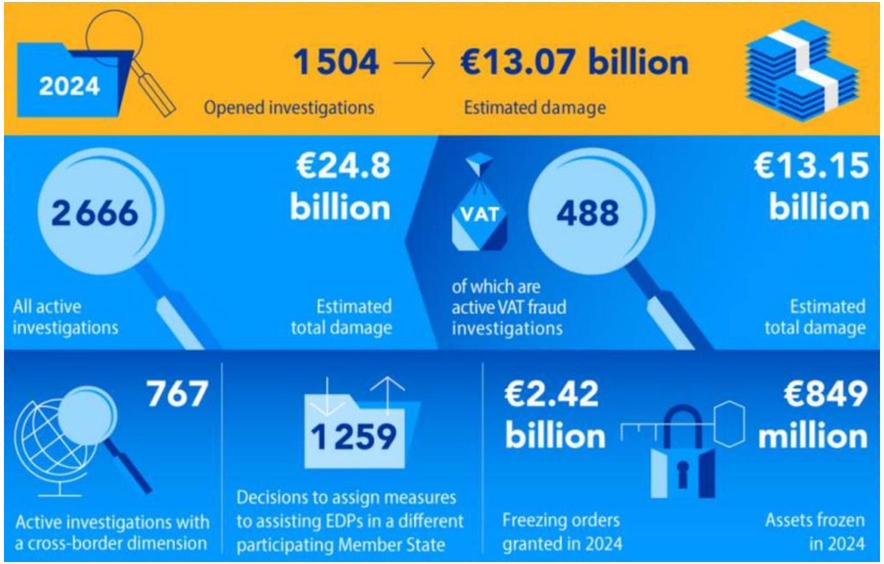
27 countries concerted effort against money laundering in Europe, Asia, North America, Colombia and Australia

1,013 arrests

2,822 banks and financial institutions supporting

10,736 fraudulent transactions

~ €32 million prevented losses



2024

1 504 → **€13.07 billion**

Opened investigations — Estimated damage

2666 All active investigations

€24.8 billion Estimated total damage

VAT

488 of which are active VAT fraud investigations

€13.15 billion Estimated total damage

767 Active investigations with a cross-border dimension

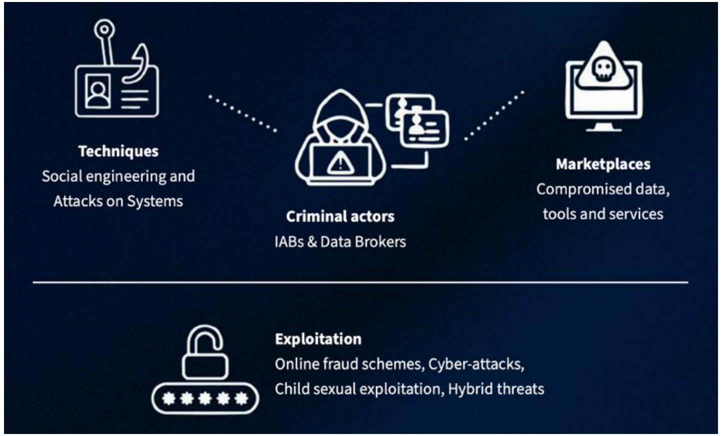
1 259 Decisions to assign measures to assisting EDPs in a different participating Member State

€2.42 billion Freezing orders granted in 2024

€849 million Assets frozen in 2024

Introduction

- EUROPOL: Millions of European citizens are **daily victims of cybercrimes**
- EPPO: cybercrime increase of 38% at the end of 2024 with damage of 24.8 billion euros
- EC3:
  - **support for Member States** in the cybercrime fight by coordinating investigations and providing technical expertise
  - **automated tools** to analyze a large amount of data and the need for specific legislation
- Cybercrime: illegal digital activity (malware and network)

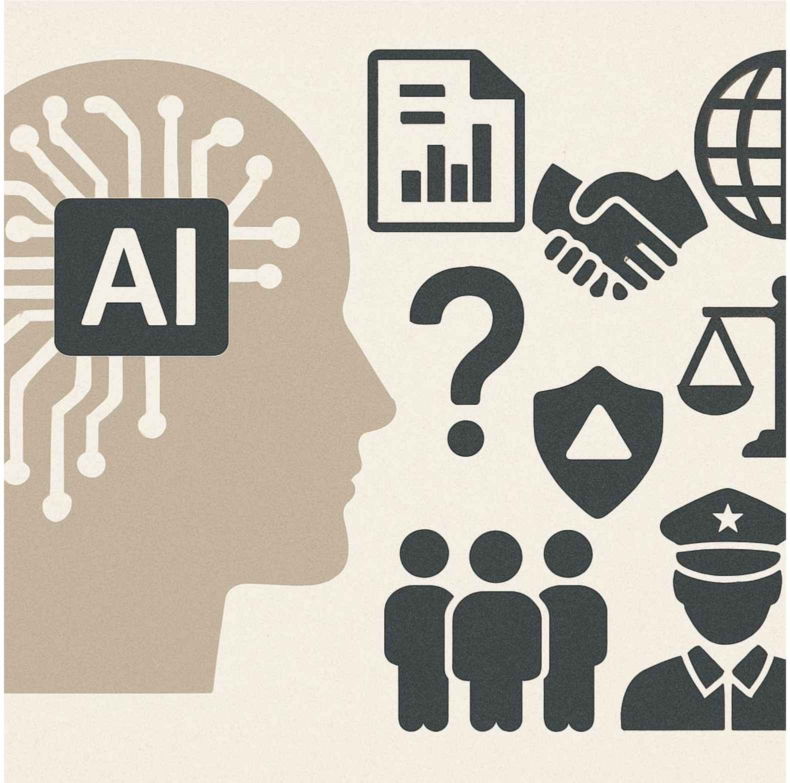

Techniques
Social engineering and Attacks on Systems
Criminal actors
IABs & Data Brokers
Marketplaces
Compromised data, tools and services
Exploitation
Online fraud schemes, Cyber-attacks, Child sexual exploitation, Hybrid threats

Introduction

- EUROPOL: Millions of European citizens are **daily victims of cybercrimes**
- EPPO: cybercrime **increase of 38% at the end of 2024** with damage of **24.8 billion euros**
- EC3:
  - **support for Member States** in the cybercrime fight by coordinating investigations and providing technical expertise
  - **automated tools** to analyze a large amount of data and the need for specific legislation
- Cybercrime: illegal digital activity (malware and network)
- Digital Forensics: digital analysis after incident
- Current techniques:
  - **Malware Analysis**: static, dynamic, hybrid, AI-based
  - **Network Intrusion Detection Systems**: fingerprint, patterns
  - **Pattern Recognition**: signature, features, ML, DL, xAI, LLMs
  - **AI in commercial and FOSS tools**: Magnet Axiom, Inseyets, Oxygen, X-Ways Forensics
  - **DF**: malware analysis, incident response, anti-forensics

Problem

- Current limitation:
  - AI models **robustness**
  - Lack of **training dataset**
  - **Real cases** reports
  - International **collaboration**
  - Lack of law **regulation**
  - DF **principles** violation
- **AI for cybercrime prevention**
- **AI can be applied in DF**:
  arrest cybercriminals, improve cybercrime detection
- AI for **privacy preservation**
- AI to prevent **investigator's shock**

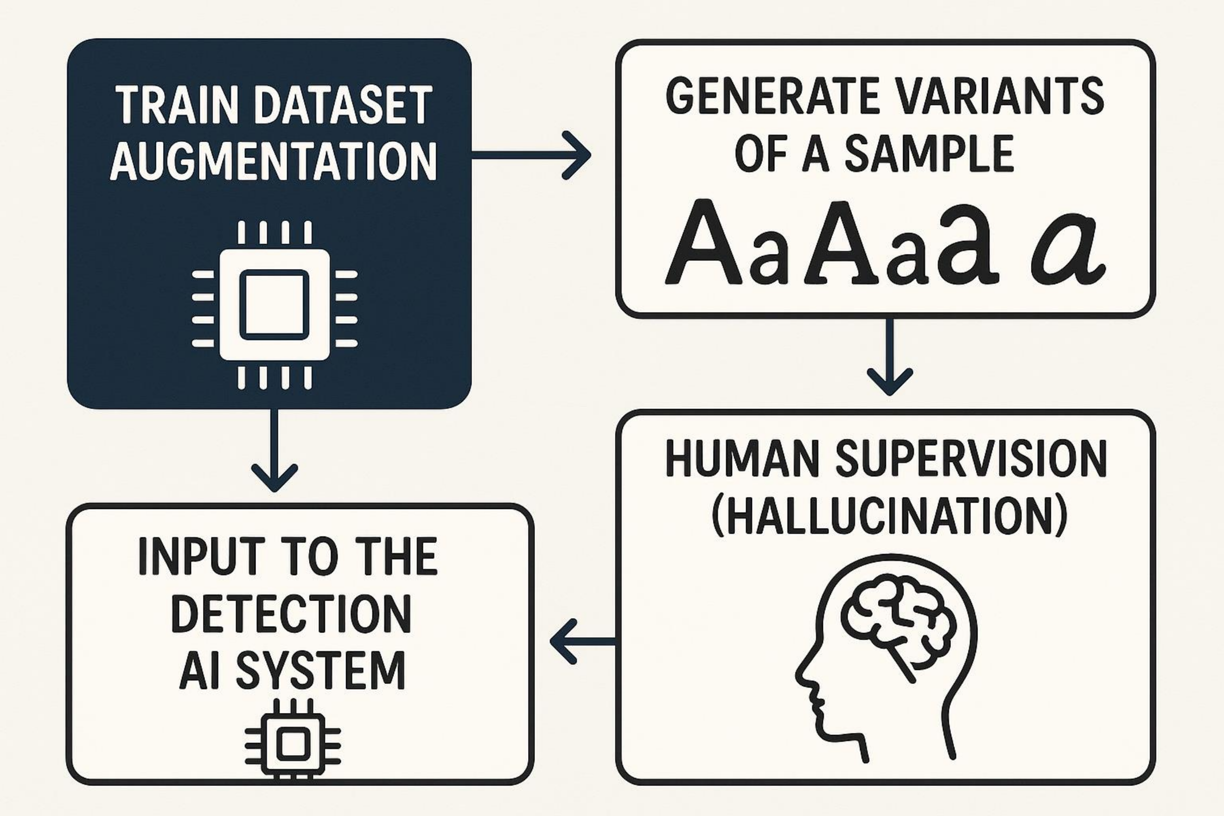


Cybercrime Detection with AI

Cybercrime Detection with AI

- AI learn **common patterns** found in attacks
- Detect **popular threat** categories according to unique features
- **Robustness**: adversarial attacks, spatial and temporal drift

Cybercrime with genAI



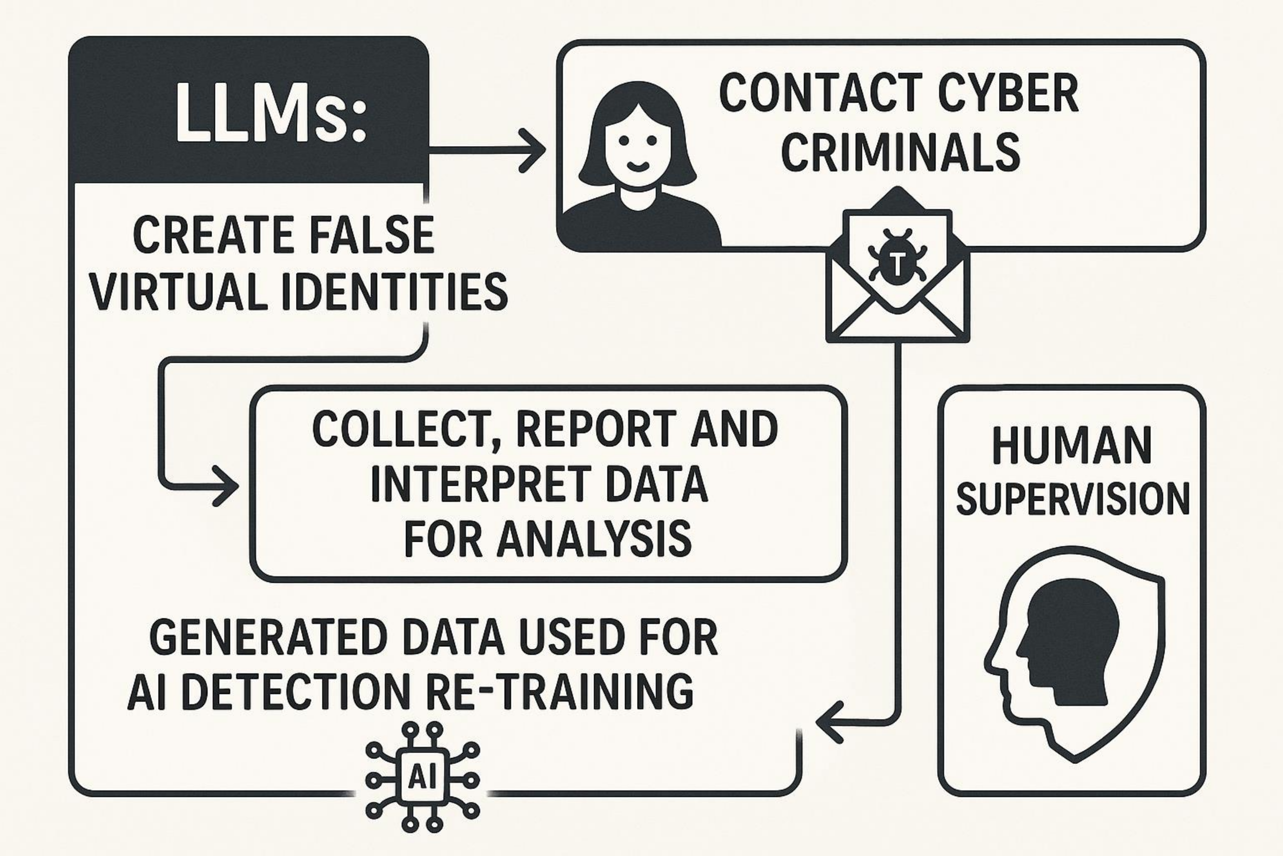
TRAIN DATASET AUGMENTATION

GENERATE VARIANTS OF A SAMPLE

HUMAN SUPERVISION (HALLUCINATION)

INPUT TO THE DETECTION AI SYSTEM

Cybercrime Detection with LLMs



LLMs:

CREATE FALSE VIRTUAL IDENTITIES

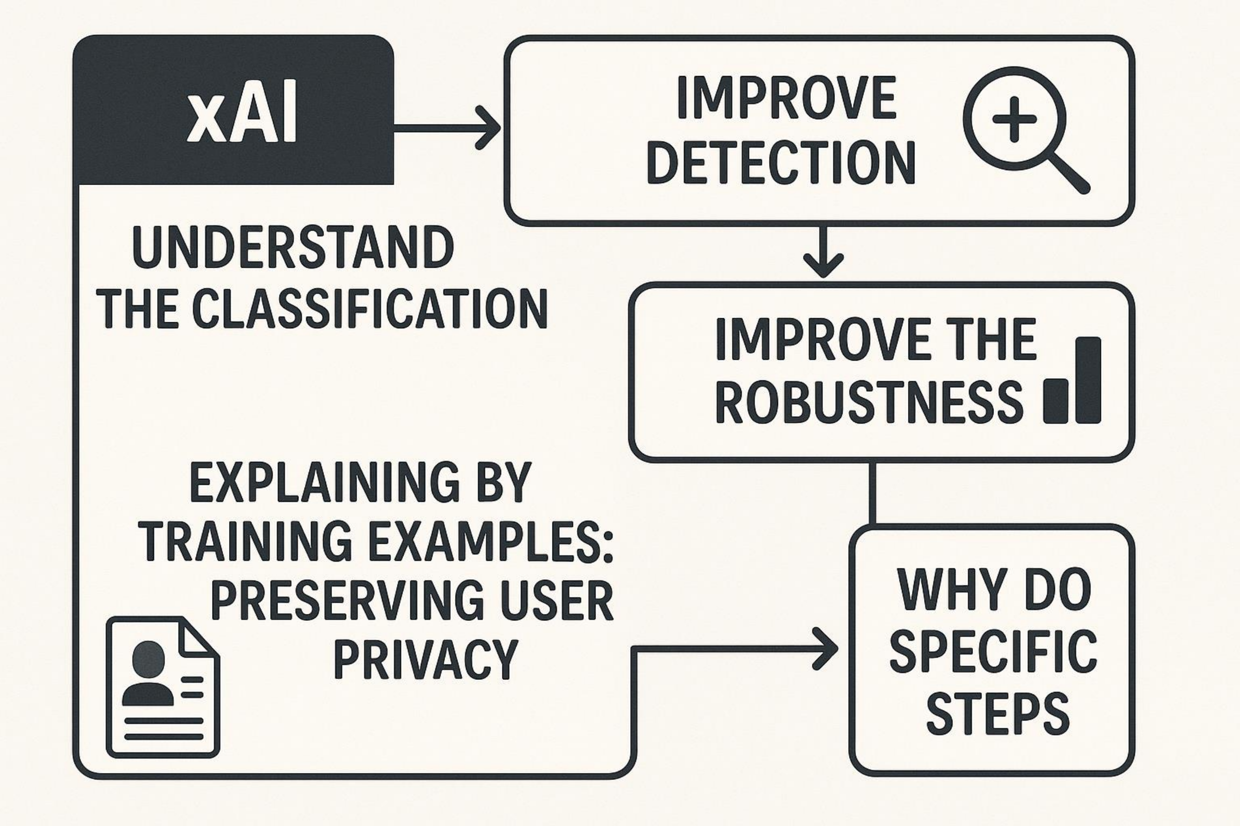
CONTACT CYBER CRIMINALS

COLLECT, REPORT AND INTERPRET DATA FOR ANALYSIS

HUMAN SUPERVISION

GENERATED DATA USED FOR AI DETECTION RE-TRAINING

Cybercrime Detection with xAI



xAI

IMPROVE DETECTION

UNDERSTAND THE CLASSIFICATION

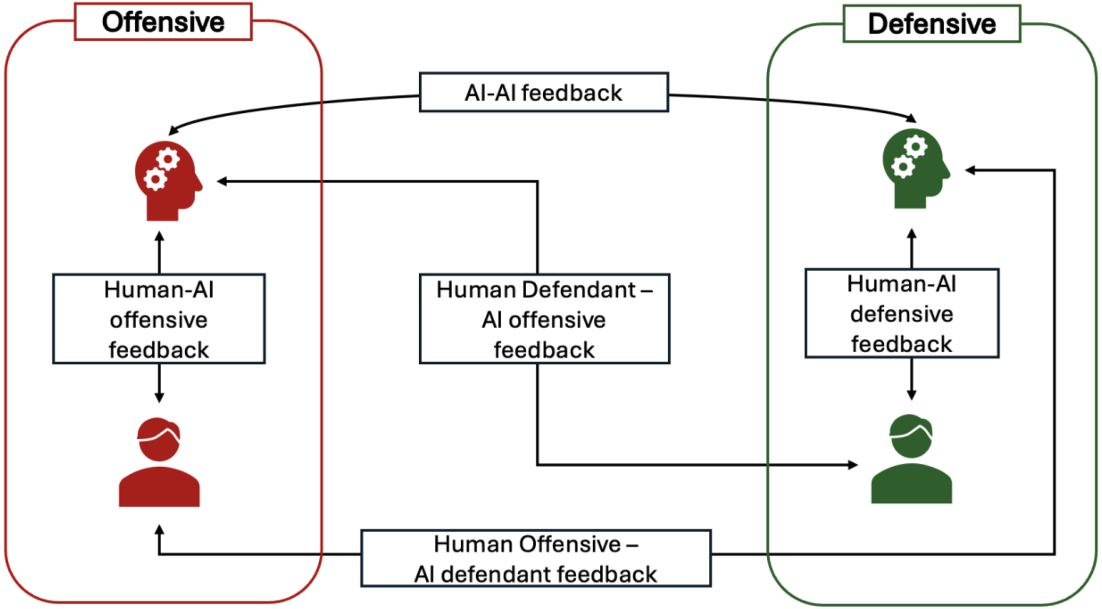
IMPROVE THE ROBUSTNESS

EXPLAINING BY TRAINING EXAMPLES: PRESERVING USER PRIVACY

WHY DO SPECIFIC STEPS

Cybercrime Detection with AI

- People are the weakest link
  - Improve skills with **games** (CTFs)
  - genAI for **new cases**
  - AI **play** against humans
  - Offensive AI for **evasive** scenarios
  - Psychology studies:
    - know **effects of attack** and **brain reaction**
    - Develop new **countermeasures**
    - Know **how the criminal think**
    - Develop **new attacks**
  - **Loop** between defensive and offensive **AI** and with **humans** feedback acting as analysts or cybercriminals


Offensive
Defensive
AI-AI feedback
Human-AI offensive feedback
Human Defendant – AI offensive feedback
Human-AI defensive feedback
Human Offensive – AI defendant feedback

Ethics and Privacy

- Police **reports** and investigations **data** for training
- Data **anonymization**
- International **collaboration**
- Non-**biased** algorithms
- **Shared database** on different world cases (federated learning)
- **Standard** AI system
- Universal **methodology**
- Need of **law** and standards
- xAI not learn by **personal information**

Digital Forensics with AI

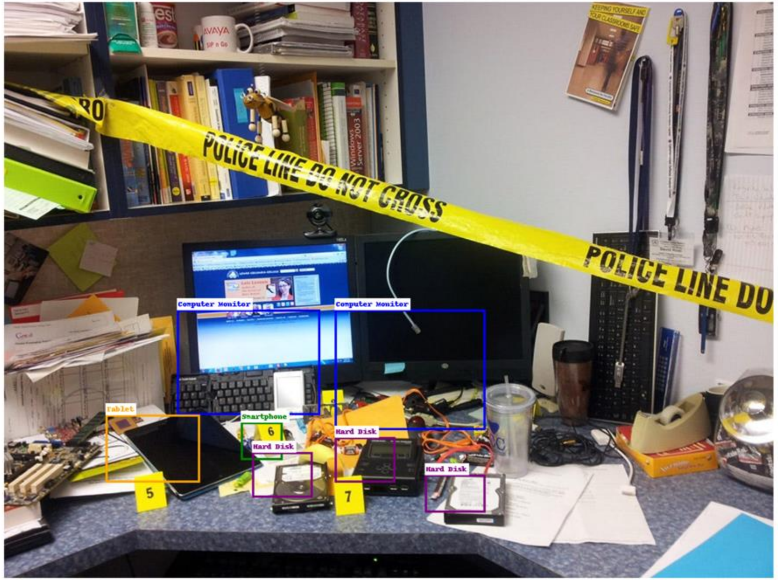
Digital Forensics with AI

- AI integrated in the 4 DF phases:
  - **Collection**: detection
  - **Examination**: acquisition
  - **Analysis**: findings
  - **Reporting**: description

- Not rely **exclusively** on AI
- Not report a result just because of AI
- **Supervision**
- AI as **additional assistant**
- xAI to **understand** model decision, **improve** robustness and classification

AI in DF Collection

- **Identify** digital devices for seizure
- Camera with **Computer Vision** (locally, no data sent to server)
- **Not forget seizure**
- Follow the **procedure**: automatically take picture, identify patterns, know state 0
- **Report** data with LLM to chain of custody



AI in DF Examination

- Digital **assistant** for the acquisition
- **Chatbot**
  - Write device specifics
  - Access database of similar cases
  - List the steps on how to acquire properly
  - Interact if something wrong
- Global **anonymized database** for training
- Technical procedure **guideline** for acquisition
- Check of **similar cases**
- Help in **corrupted data** (restore, repair)

Environment Preparation

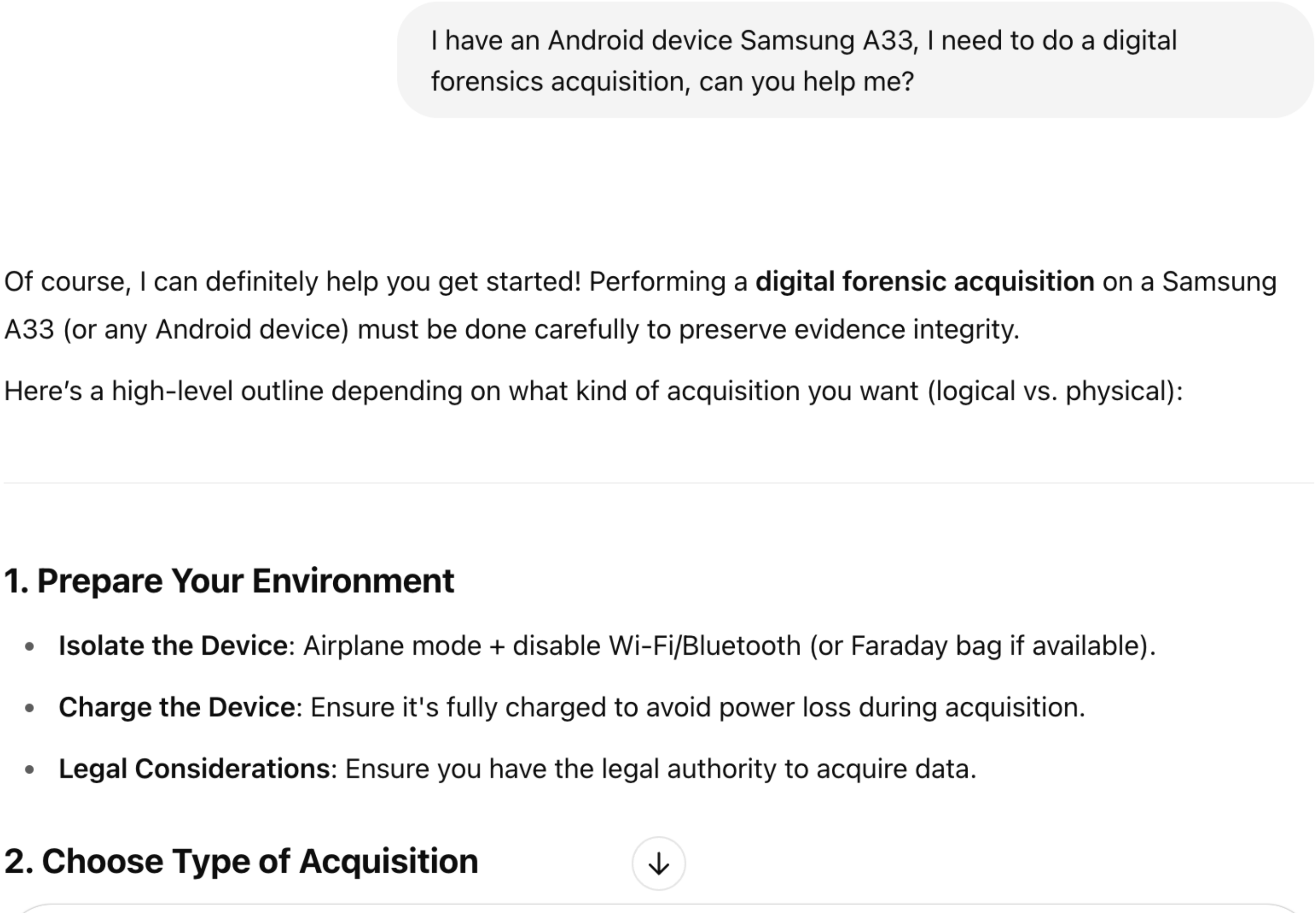
Choose Acquisition Type

Select Tools

Detailed Steps

Extract Data

Preserve Acquisition



I have an Android device Samsung A33, I need to do a digital forensics acquisition, can you help me?

Of course, I can definitely help you get started! Performing a **digital forensic acquisition** on a Samsung A33 (or any Android device) must be done carefully to preserve evidence integrity.

Here’s a high-level outline depending on what kind of acquisition you want (logical vs. physical):

---

1. Prepare Your Environment

- **Isolate the Device**: Airplane mode + disable Wi-Fi/Bluetooth (or Faraday bag if available).
- **Charge the Device**: Ensure it's fully charged to avoid power loss during acquisition.
- **Legal Considerations**: Ensure you have the legal authority to acquire data.

2. Choose Type of Acquisition

2. Choose Type of Acquisition

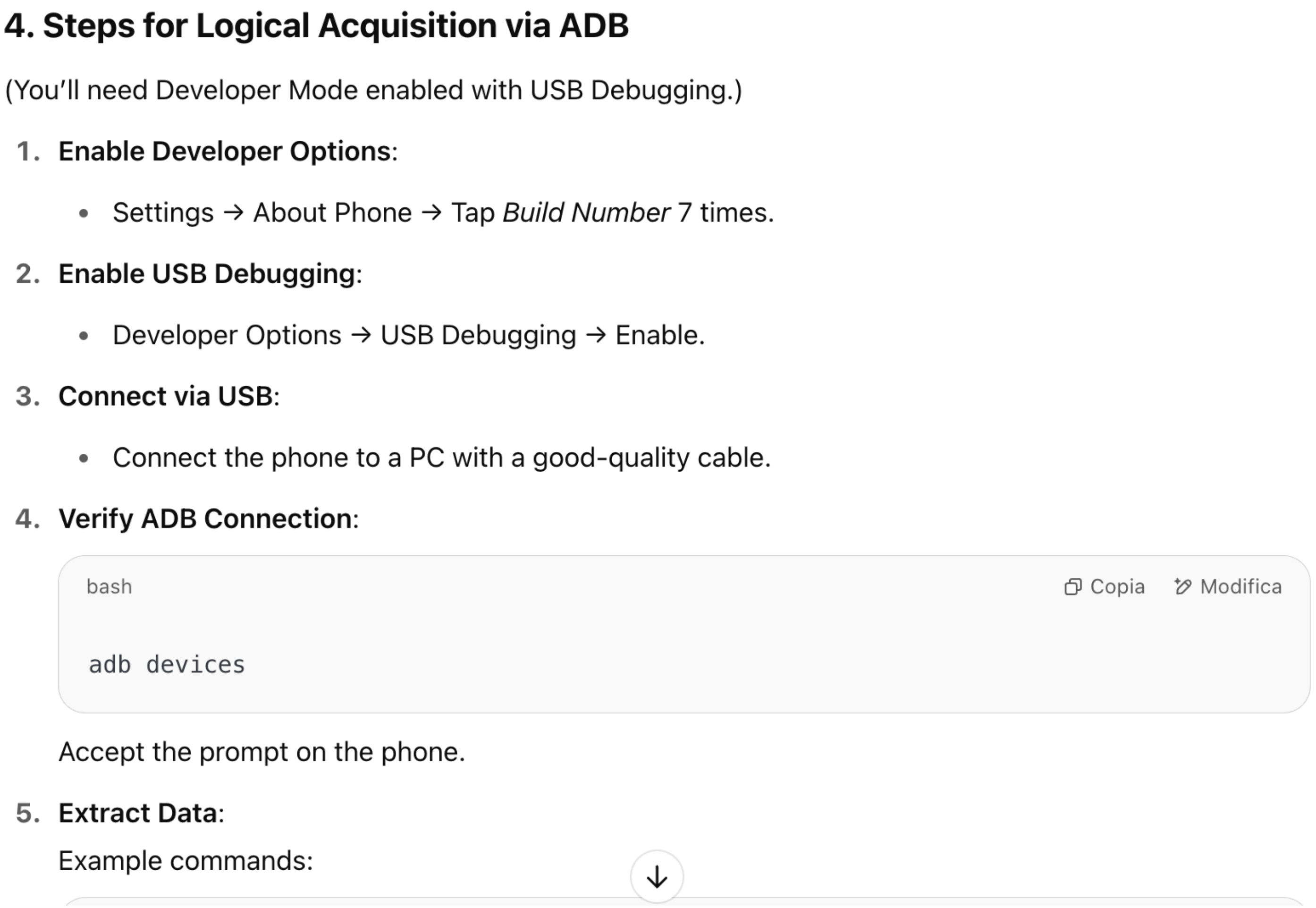
- **Logical Acquisition**: Extract accessible files (contacts, SMS, media, app data).
- **Physical Acquisition**: Create a bit-by-bit copy (more complex on newer phones due to encryption).

For Samsung devices running Android 11/12/13, full **physical acquisition** without root access is extremely difficult because of **Full-Disk Encryption (FDE)** and security policies.

---

3. Tools You Can Use

- **ADB (Android Debug Bridge)**: Basic logical acquisition.
- **Magnet AXIOM, Cellebrite UFED, Oxygen Forensics**: Commercial tools for more advanced acquisition (if you have access).
- **Open-source tools**:
    - **Autopsy + Android Analyzer Modules**
    - **ADB + `dd` (if rooted)**
    - **AFLogical OSE (Logical acquisition via A**



4. Steps for Logical Acquisition via ADB

(You'll need Developer Mode enabled with USB Debugging.)

1. **Enable Developer Options**:
   - Settings → About Phone → Tap *Build Number* 7 times.
2. **Enable USB Debugging**:
   - Developer Options → USB Debugging → Enable.
3. **Connect via USB**:
   - Connect the phone to a PC with a good-quality cable.
4. **Verify ADB Connection**:

```bash
adb devices
```

   Accept the prompt on the phone.

5. **Extract Data**:

   Example commands:

5. **Extract Data**:

   Example commands:

   ```bash
   adb pull /sdcard/ /path/to/your/pc/sdcard_backup/
   adb backup -apk -shared -all -f backup.ab
   ```

   (Backup file will be in `.ab` format — can be converted with `abe.jar` to a tar file.)

---

5. Considerations for Physical Acquisition

Physical acquisition is much more complex:

- **Custom Recovery (TWRP)** + `dd` command for block-level imaging (needs an unlocked bootloader → risky and voids warranty).
- **Chip-off or JTAG** (advanced lab hardware needed).

On a stock A33:

- Full physical acquisition is *extremely unlikely* w ut advanced equipment or vendor cooperation.

6. Preserving Evidence

- Create a **hash** (MD5/SHA256) of acquired files/images.

  Example:

  ```bash
  sha256sum backup.ab
  ```

- Maintain a **Chain of Custody** document — who handled the device, when, and how.

---

7. Example for ADB Backup

```bash
adb backup -apk -shared -all -f samsung_a33_backup.ab
```

Then you can convert it to TAR:

Important Limitations

- New Android versions heavily encrypt data — without root or custom recovery, full data acquisition is limited.
- Some apps (like WhatsApp) have end-to-end encrypted backups.
- Samsung's Knox security can prevent access to certain areas (e.g., Secure Folder).

Would you like me to guide you through a **live logical acquisition** with ADB? Or are you aiming for **full physical acquisition** (e.g., using more advanced methods like TWRP or chip-off)? Let's pick based on your goal and resources! 🚀

Do you have access to forensic software like Cellebrite or Magnet AXIOM?

how to root the device? do I loose data?

Good questions — rooting is a serious step, especially for forensic work. Let's break it down carefully:

Can You Root the Samsung A33?

Yes, but **Samsung Galaxy A33** (especially models with Android 12/13/14) has **Knox security** and **bootloader lock**, which make rooting tricky and risky:

1. **Bootloader Unlocking is required** before rooting.
2. **Rooting will trip Knox**, and you'll **lose all existing data** during bootloader unlock.

Will You Lose Data?

✅ **Yes**, unlocking the bootloader **forces a factory reset** for security reasons.
➡️ **All data will be wiped**: photos, messages, app data, everything.

That's why **rooting is *not ideal* for forensic acquisition**, because:

- It **destroys** evidence already on the device (wiping data = loss of evidence integrity).
- **Knox e-fuse** will be tripped permanently (can't revert).
- Tripping Knox can disable services like Secure Folder and Samsung Pay.
- The device will show "Bootloader unlocked" warnings — making it *obvious* that the device was modified.

Basic Rooting Steps for Samsung A33 (Theoretical)

(⚠️ Not recommended for forensics because of data wipe)

1. **OEM Unlock**:
    - Developer Options → Enable *OEM Unlock*.
2. **Unlock Bootloader**:
    - Power off device.
    - Press Volume Up + Down → Connect to PC via USB → Enter Download Mode.
    - Long-press Volume Up to unlock the bootloader.
    - ⚠️ **This wipes all data.**
3. **Flash a Custom Recovery (e.g., TWRP)**:
    - Odin tool (on PC) to flash TWRP.
4. **Root the Device**:
    - Flash Magisk via TWRP to gain root access.

⚠️ Forensics Recommendation

Do NOT root if your goal is forensic preservation of data.
Because:

- **Data wipe** = loss of evidence.
- **Knox trip** = permanent hardware change.
- **Legal evidence chain** could be broken.

Alternative: Non-Root Logical Acquisition

If you can get into the device with the screen unlocked and USB Debugging enabled:

- Use **ADB** for logical backup.
- Use **forensic tools** like Magnet AXIOM or Oxygen Forensic Detective (they work with screen unlock).
- Some tools support **Android Agent-based extraction** without root.

If the device is locked and you cannot enable USB Debugging or unlock the screen, you'll need **advanced methods**:

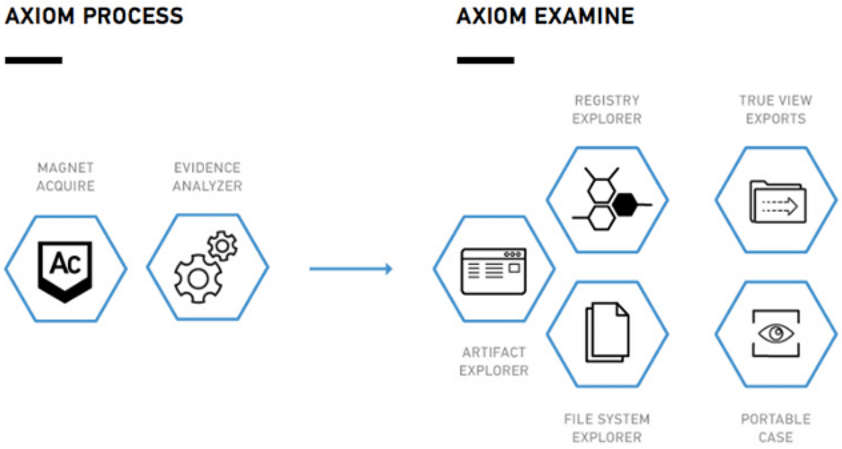
- **Chip-off** (physical memory extraction).
- **ISP (In-System Programming)** techniques.
- **Vendor Assistance** (if you have proper authority and chain of custody).

Summary

Option	Data Loss?	Good for Forensics?	Comments
Rooting & Unlocking Bootloader	✅ Wipes Data	❌ No	Not good, you lose evidence.
Logical Acquisition (ADB)	❌ No	✅ Yes	Best if screen is unlocked.
Advanced Physical Methods	❌ No	✅ Yes	Very complex, hardware needed.

AI in DF Analysis

- Classify **multimedia** files
- Improved in terms of **robustness** and **accuracy**
- Train AI on **anti-forensics** and anti-analysis
- Detect **hidden patterns** in data
- Help retrieve hidden or **damaged content**
- LLM to **report** the findings
- Global **anonymized database** of similar cases



AXIOM PROCESS

MAGNET ACQUIRE

EVIDENCE ANALYZER

AXIOM EXAMINE

REGISTRY EXPLORER

TRUE VIEW EXPORTS

ARTIFACT EXPLORER

FILE SYSTEM EXPLORER

PORTABLE CASE

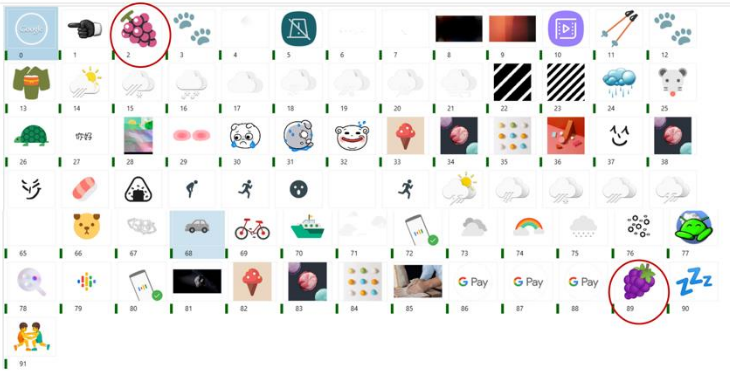
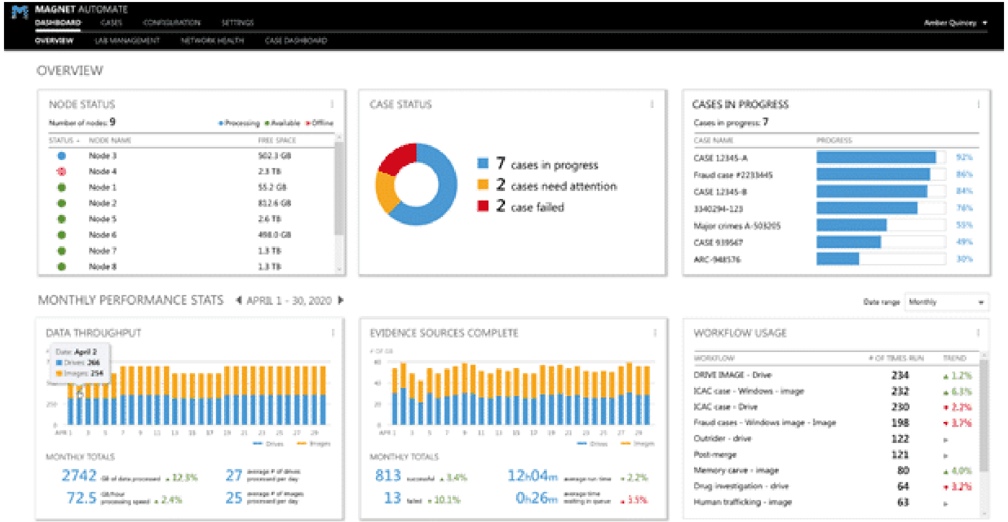


Figure 5: The image shows the pictures detected as drugs but that clearly are not drugs (false positive rates) such as the grape (red circled) for the drug detection in chat done by Magnet AI tool.

S. Sanna, L. Regano, D. Maiorca, G. Giacinto. (2024). Exploring the Robustness of AI-Driven Tools in Digital Forensics: A Preliminary Study. 10.48550/arXiv.2412.01363

AI in DF Reporting

- **Summary findings** written with NLP and LLMs
- Important for the **background report** section and glossary
- Explain **technical concepts** to unexpert people (judge, lawyer, etc)
- Read **external sources** for bibliography if needed
- **Summarize** the collected documents in different phases
- NLP and LLMs to **evaluate** DF's law **understandability**

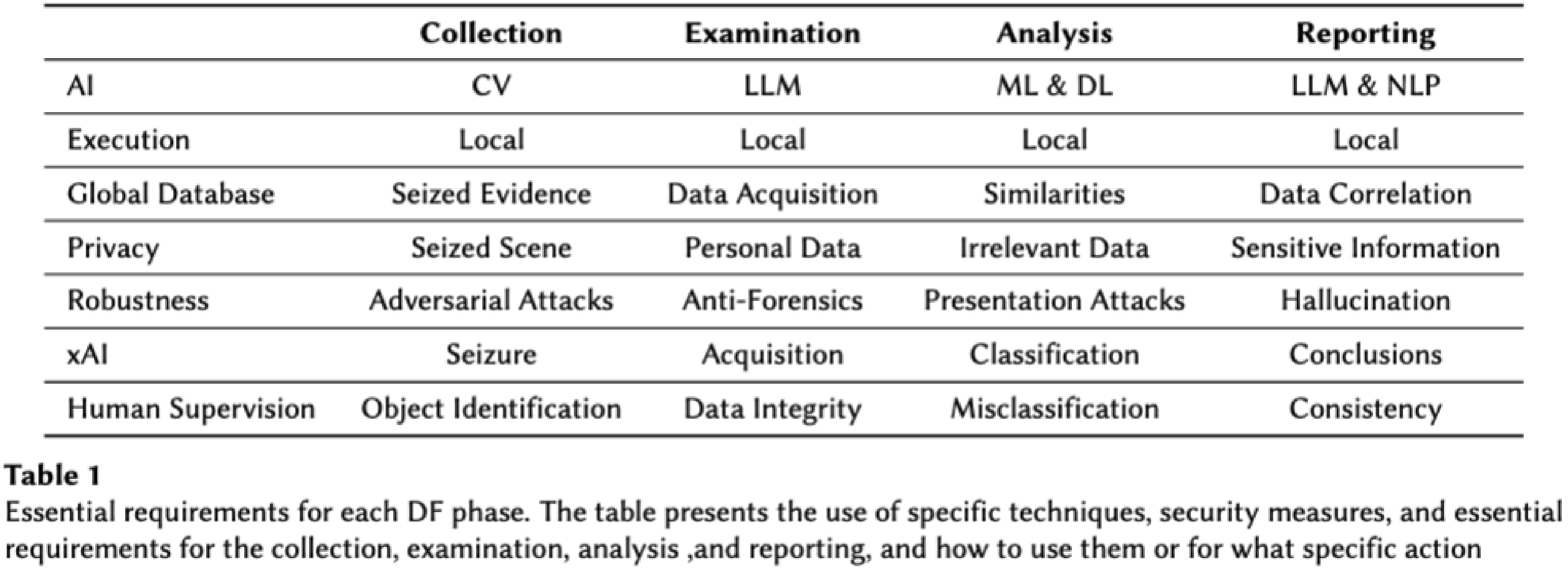


STATUS	NODE NAME	FREE SPACE
●	Node 3	502.3 GB
●	Node 4	2.3 TB
●	Node 1	55.2 GB
●	Node 2	812.6 GB
●	Node 5	2.6 TB
●	Node 6	498.0 GB
●	Node 7	1.3 TB
●	Node 8	1.3 TB

CASE NAME	PROGRESS
CASE 12345-A	92%
Fraud case #2233445	86%
CASE 12345-B	84%
3340294-123	76%
Major crimes A-503205	55%
CASE 939567	49%
ARC-948576	30%

WORKFLOW	# OF TIMES RUN	TREND
DRIVE IMAGE - Drive	234	▲ 1.2%
ICAC case - Windows - image	232	▲ 6.3%
ICAC case - Drive	230	▼ 2.2%
Fraud cases - Windows image - Image	198	▼ 3.7%
Outrider - drive	122	▸
Post-merge	121	▸
Memory carve - image	80	▲ 4.0%
Drug investigation - drive	64	▼ 3.2%
Human trafficking - image	63	▸

AI in DF



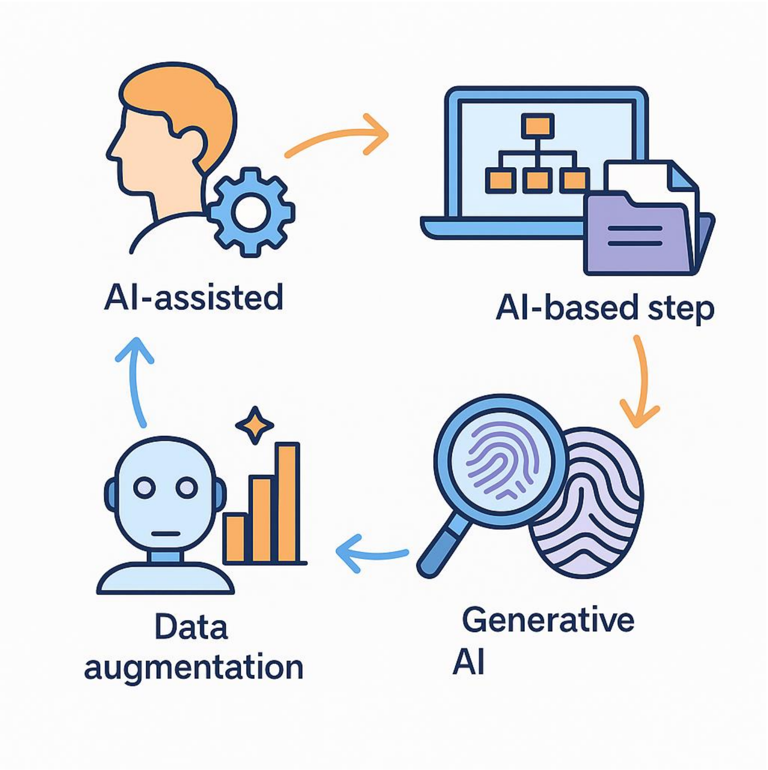
	Collection	Examination	Analysis	Reporting
AI	CV	LLM	ML & DL	LLM & NLP
Execution	Local	Local	Local	Local
Global Database	Seized Evidence	Data Acquisition	Similarities	Data Correlation
Privacy	Seized Scene	Personal Data	Irrelevant Data	Sensitive Information
Robustness	Adversarial Attacks	Anti-Forensics	Presentation Attacks	Hallucination
xAI	Seizure	Acquisition	Classification	Conclusions
Human Supervision	Object Identification	Data Integrity	Misclassification	Consistency

Table 1

Essential requirements for each DF phase. The table presents the use of specific techniques, security measures, and essential requirements for the collection, examination, analysis ,and reporting, and how to use them or for what specific action

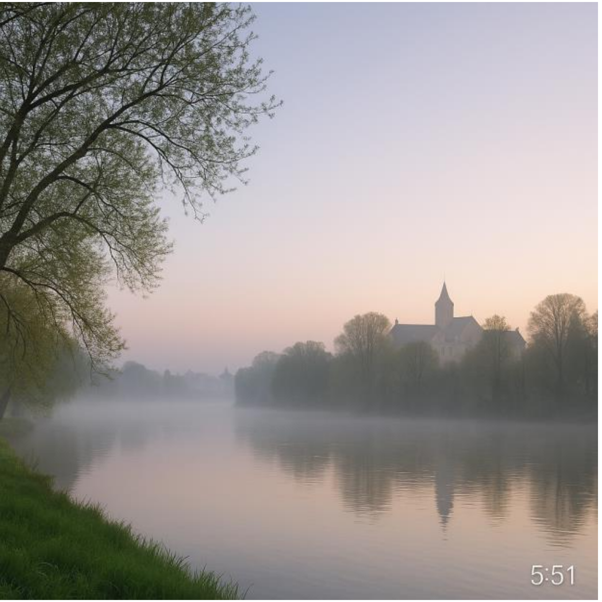
AI helping DF

- **AI-assisted** consultancies
- **Common patterns** in worldwide cases
- **Integration** to human analysis
- **Pre-analysis** as a starting point
- Challenging scenarios (**anti-forensics**)
- AI-assisted: 4 phases
- **genAI** for data augmentation, new cases, training



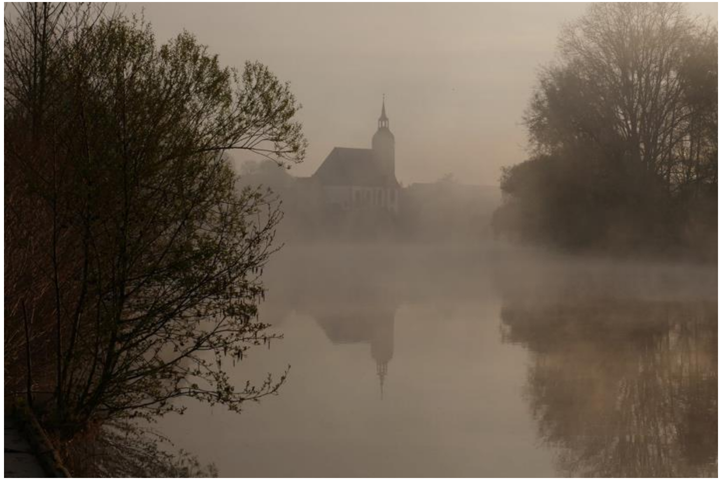
DF Analysis of genAI data

- Increasing spread of genAI data, generated for free and easily



Generated

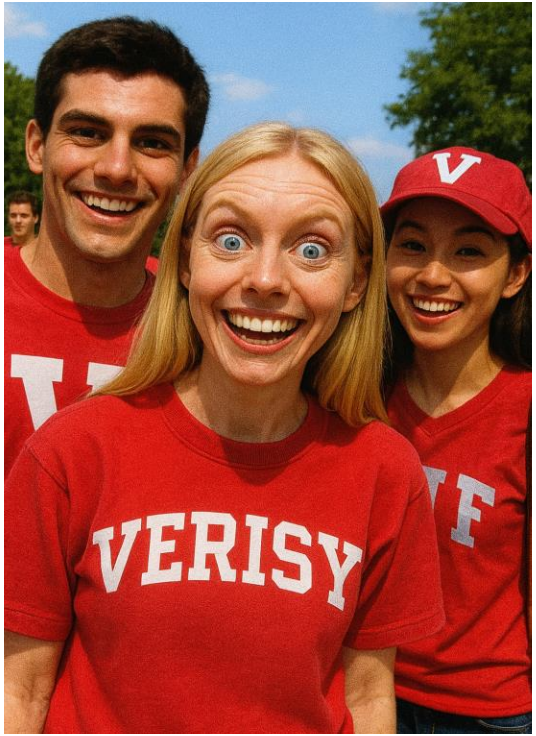
Painting: https://tinyurl.com/3uutbkvv



Real (similar to generated): https://www.freizeitbad-riff.de/wettbewerb/122-morgenstunde-2

DF Analysis of genAI data

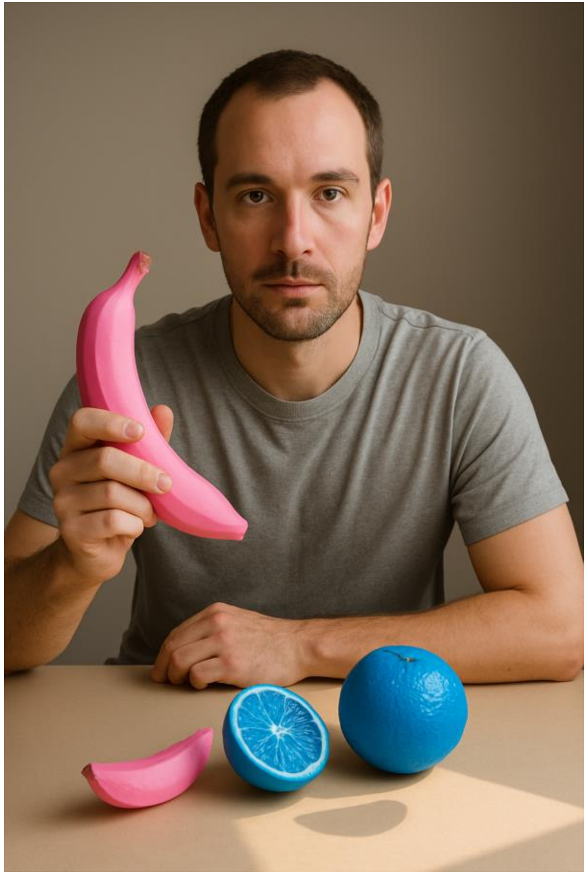
- *How to prove that the data is generated by AI mimicking the real person?*
- *How to prove who made the synthetic audio/video/image?*



Unreal shadow in openair (left to right for boy and blonde girl, right to left for the girl with the hat)

DF Analysis of genAI data

- **Steganography**: watermark identifying the used tool
- Deep **file structure** analysis
- Comparison with **real similar data**
- Comparison with **generated similar data** with literature tools
- Recurrent **unrealistic pattern**, file structure used by AI to detect genAI data



Case Study:
LLMs Chatbots for Steganography

LLMs Chatbots for Steganography

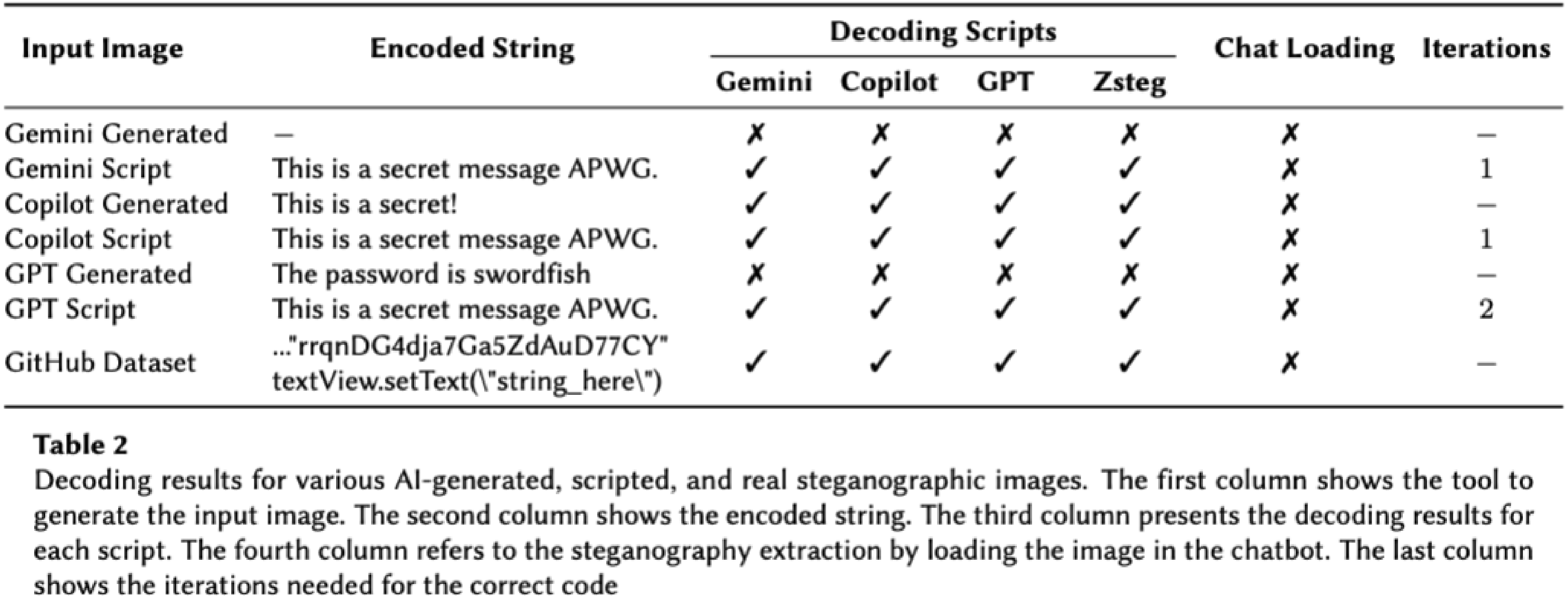
- Gemini 2.5, Copilot, chatGPT-4o
- *Can you generate an image with a hidden secret with steganography techniques? I also want the script python to decode it.*
- png images: no decoded string, *zsteg* empty
- *Generate a python script to encode a given message with a given image.*
- Successful encoding and decoding with synthetic and real data

Gemini

Copilot

ChatGPT

LLMs Chatbots for Steganography



Input Image	Encoded String	Decoding Scripts				Chat Loading	Iterations
		Gemini	Copilot	GPT	Zsteg		
Gemini Generated	–	✗	✗	✗	✗	✗	–
Gemini Script	This is a secret message APWG.	✓	✓	✓	✓	✗	1
Copilot Generated	This is a secret!	✓	✓	✓	✓	✗	–
Copilot Script	This is a secret message APWG.	✓	✓	✓	✓	✗	1
GPT Generated	The password is swordfish	✗	✗	✗	✗	✗	–
GPT Script	This is a secret message APWG.	✓	✓	✓	✓	✗	2
GitHub Dataset	..."rrqnDG4dja7Ga5ZdAuD77CY" textView.setText(\"string_here\")	✓	✓	✓	✓	✗	–

Table 2
Decoding results for various AI-generated, scripted, and real steganographic images. The first column shows the tool to generate the input image. The second column shows the encoded string. The third column presents the decoding results for each script. The fourth column refers to the steganography extraction by loading the image in the chatbot. The last column shows the iterations needed for the correct code

https://github.com/slsanna/LLMs-Chatbots-for-Steganography

Conclusions

- Current methodologies limitations
- How AI can improve cybercrime detection
- How AI can help in DF
- Forensically analyse genAI data
- Easy use of AI as defence and attack
- Future studies:
  - Improve performances
  - Development of ad-hoc tools
  - Consider anti-forensics scenarios
  - AI-vs-AI for defense and attack

sAIfer Lab

Joint lab on Safety and Security of AI

Thank you for your attention

Silvia Lucia Sanna

silvial.sanna@unica.it

www.saiferlab.ai

UNICA UNIVERSITÀ DEGLI STUDI DI CAGLIARI

Università di Genova