dfki ai — Deutsches Forschungszentrum für Künstliche Intelligenz / German Research Center for Artificial Intelligence

Dr. Tobias Wirth[a], Andrey Guzhov[a], Dr. Matthias Müller[b], Tobias Fischer[b], Lucas Howes[c], Dominik Ospelt[c]

[a] German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
[b] State Criminal Police Office of Rhineland-Palatinate, Mainz, Germany
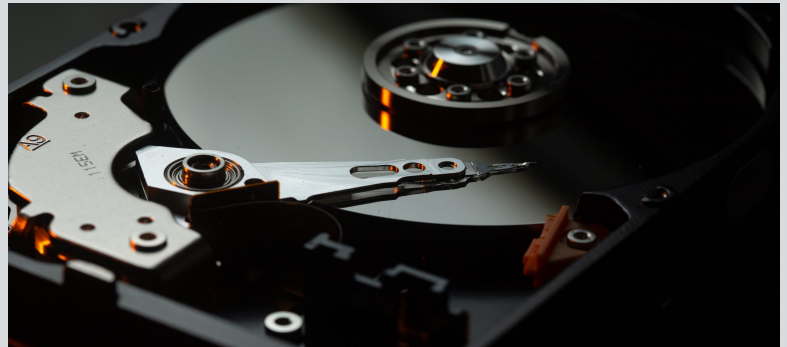[c] Binary Impact GmbH, Niederelbert, Germany

Carve-DL

AI Combats Cybercrime

Deep Learning for Digital Forensics

Data and Challenges

Source: SWR
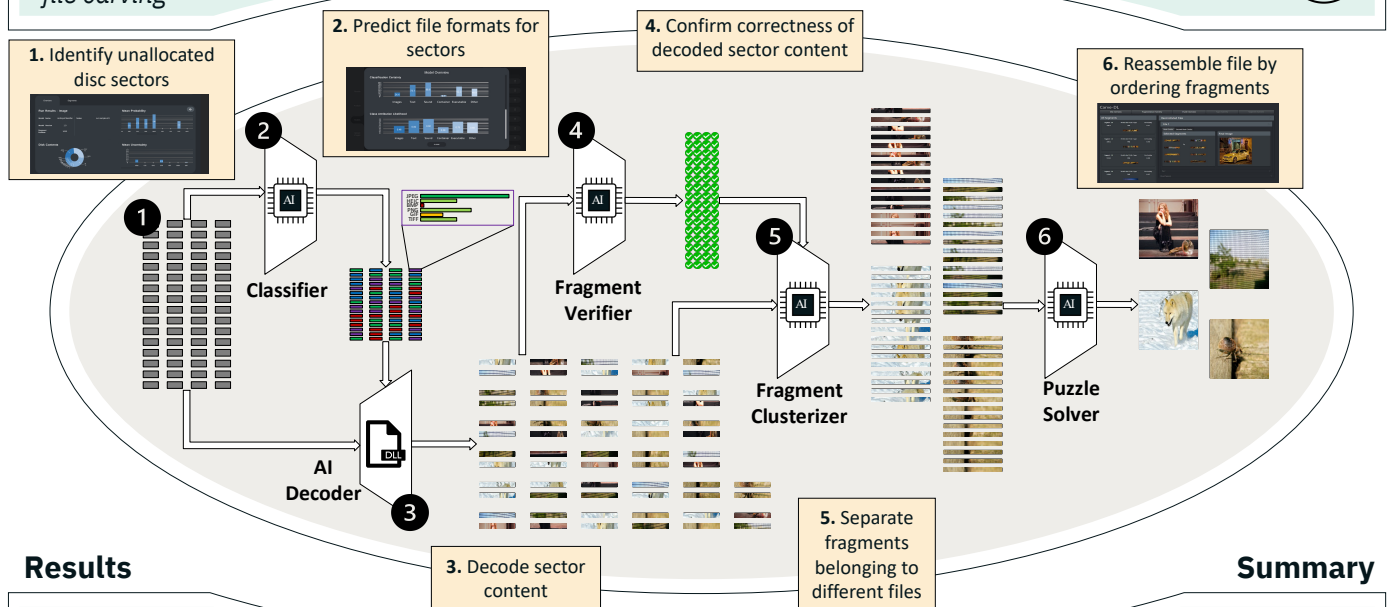
Underground cybercrime data center **"Cyber Bunker"** was seized in 2019 by German police

Hard disks containing **>300 TB** of partially fragmented or corrupted data were found

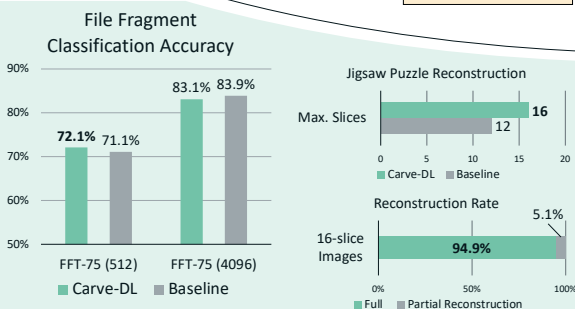**Carve-DL** Use Case → *Digital evidence restoration by AI-aided file carving*

Project Goals

- Build state-of-the-art data reconstruction pipeline with deep learning techniques
- Develop intuitive user interface for forensic investigators
- Employ hardware acceleration to facilitate processing of massive amounts of data

Pipeline

1. Identify unallocated disc sectors

2. Predict file formats for sectors

4. Confirm correctness of decoded sector content

6. Reassemble file by ordering fragments

1 → **Classifier** (2) → **AI Decoder** (3) → **Fragment Verifier** (4) → **Fragment Clusterizer** (5) → **Puzzle Solver** (6)

3. Decode sector content

5. Separate fragments belonging to different files

Results

File Fragment Classification Accuracy

	Carve-DL	Baseline
FFT-75 (512)	72.1%	71.1%
FFT-75 (4096)	83.1%	83.9%

Jigsaw Puzzle Reconstruction

Max. Slices — Carve-DL: 16, Baseline: 12

Reconstruction Rate

16-slice Images — Full: 94.9%, Partial Reconstruction: 5.1%

Summary

In progress...

- ✓ Data and architecture requirements analyzed
- ✓ Synthetic and real-world datasets for model training and validation generated
- ✓ AI and UI/UX workflow designed
- ✓ Deep Learning models developed
  - ✓ Best classification accuracy on FFT-75 (512) → **72.1%**
  - ✓ Largest number of image slices → **16**
  - ✓ Highest perfect reconstruction ratio → **94.9%**