Advanced Seminar on DeepFake Detection

C. Rathgeb and B. Chopin

da/sec - Biometrics and Security Research Group, Hochschule Darmstadt

Darmstadt, 20.10.2026

Contact and Website

- Supervisors: C. Rathgeb and B. Chopin
- Email addresses:
  `christian.rathgeb@h-da.de`
  `baptiste.chopin@h-da.de`
- Website:
  `https://dasec.h-da.de/`
- Seminar website:
  `https://dasec.h-da.de/advanced-seminar-masterseminar-ws-2026-27/`

Procedure

- Students select preferred topics and groups (up to 2 students depending on the topic)
- Required materials will be handed over by the supervisors: databases, software, evaluation scripts etc.
- **Progress reports** have to be sent to the supervisor(s) before the following dates:
  1) 29th November
  2) 17th January
- Additional appointments shall be arranged individually and on demand.

Procedure (cont'd)

- Each group has to prepare a term paper (6-8 pages) using the IEEE conference template: `https://www.ieee.org/conferences/publishing/templates.html`
- The term paper must be submitted by **1st March**
- Structure: Introduction, related work (one third), method, evaluation
- Figures, tables and 15-30 Citations
- The final presentation will be 40 minutes per group + 20 minutes discussion of the results
- Suggested dates for the final presentations: **mid March** (date will be chosen via poll)
- A grade based on the term paper and final presentation

Overview

- LLM based deepfake detection
- Breaking deepfake watermarks
- Foundation models for zero shot deepfake detection
- Counting subjects for deepfake detection

The CCDF dataset

CCDF is a dataset containing AI-generated video reproducing the content and format or real CCTV footage depicting crimes and accidents. Data representative of real malicious AI-generated content found online.

- Generated using 3 state-of-the art commercial models: Veo3.1, Grokk imagine, Sora2
- Modern deepfake detectors both image- and video-based are unable to reliably detect these deepfakes.
- Topics are related to this dataset

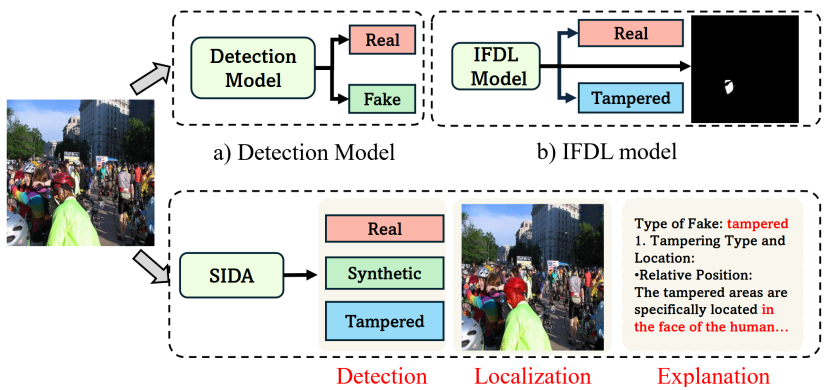


LLM based deepfake detection

Large Language Models (LLMs) and Vision Language Models (VLMs) have been adapted to build powerful deepfake detectors. However we do not knows how well they generalize to other generators.

Tasks: Search and implement models for video (and image) deepfake detection that rely on Large Language Models (LLMs).

- Implementation of LLm-based deepfake methods on our server infrastructure.
- Documentation: Term paper, presentation
- Handout: initial list of relevant approaches will be provided, evaluation metrics



Breaking deepfake watermarks

Recent generative models embed hidden watermark to ensure that images and video are not misused. While these watermarks are claimed to be extremely resistant, how robust are they really?
Tasks: Try to break synthid watermarking from video generated with gemini models

- Implement and evaluate robustness test including several image processing pipelines
- Documentation: Term paper, presentation
- Handout: initial set of videos, literature



Foundation models for zero shot deepfake detection

Foundation models for videos are now able to perform various task like video understanding with impressive performance. How well do these foundations models perform when detecting video deepfakes Tasks: Find and benchmark foundation models on deepfake detection (image or video).

- Implementation and evaluation of foundation models for deepfake detection on our server infrastructure
- Documentation: Term paper, presentation
- Handout: Initial list of relevant foundation models, deepfake database, evaluation metrics



Counting subjects for deepfake detection

The more complex an action in a generated video is, the more likely artifacts appear. Therefore it is of interest to evaluate detection performance of deepfake detectors in relation to the number of subjects present in a video/image.

Tasks: Correlate the number of subjects in videos with accuracy of deepfake detectors.

- Implementation of person detector to annotate (deepfake) videos.
- Documentation: Term paper, presentation
- Handout: Deepfake database, detector scores, evaluation metrics



Select your own topic

- Students are also invited to proposed their own topic!

Topic selection process

- Send your topic choice to `christian.rathgeb@h-da.de` by the end of the week
- First come, first serve – so you might choose more than one topic (1st and 2nd choice)