



Politecnico
di Torino
SmartData@PoliTO



June 5th, 2025, 5:00 PM CEST

SmartTalk: Covivio, Sala Grande

<https://smartdata.polito.it/category/smarttalks/>



Kai Huang

Kai is a 2nd year Ph.D. student at DAUIN. He obtained the M.Sc. in ICT for Smart Societies at Politecnico di Torino, and he is member of the SmartData center. His research interest lies at the intersection of data science and network measurements.

CFA-Bench: Cybersecurity Forensic LLM Agent Benchmark and Testing

ABSTRACT

The rapid evolution of cyber threats and the increasing sophistication of cyberattacks have made **digital forensics** a cornerstone of modern cybersecurity. In this context, Large Language Models (LLMs) present a compelling opportunity to augment cybersecurity forensics.

To investigate the capabilities of LLM agents in performing cybersecurity forensic tasks and enable fair comparisons, we introduce **CFA-bench**, a novel benchmark incorporating real-world incident response scenarios to evaluate their forensic reasoning abilities. We conduct tests on different LLM agent architectures. While preliminary, our findings demonstrate the **potential of LLM agents in cybersecurity forensics**, revealing their strengths and critical areas for improvement. We are continuing to enrich the benchmark with broader scenarios and polish its structure, while also refining agent design to improve their performance on complex cybersecurity forensic tasks.



Season 5