Fachhochschule Kiel
University of Applied Sciences

Graph of Effort

Quantifying Risk of AI Usage for Vulnerability Assessment

Anket Mehra Malte Prieß Andreas Aßmuth

8th April 2025

Kiel University of Applied Sciences

Anket Mehra

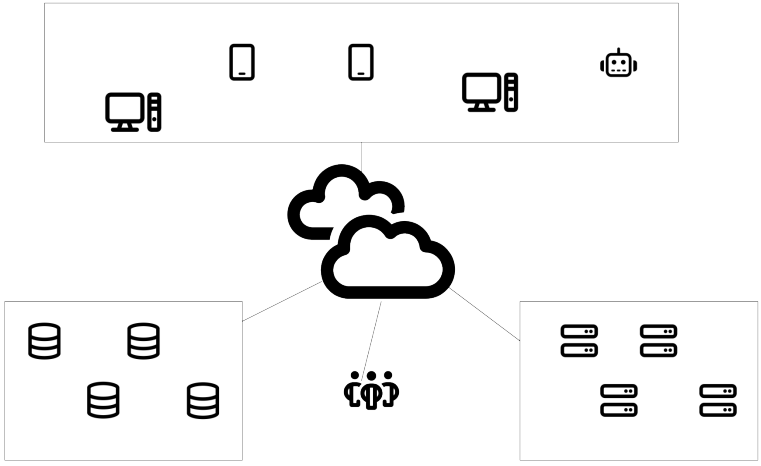
- Background
  - Masterstudent Computer Science
  - Software Developer
  - Data Scientist
- Publications
  - Improving Applicability of Deep Learning based Token Classification models during Training, ArXiv Preprint, 2025
  - Evaluierung des Dense Passage Retrievals zur Dokumentensuche in Behörden im Vergleich zu BM25, AKWI, 2022

Contact: anket.mehra@student.fh-kiel.de

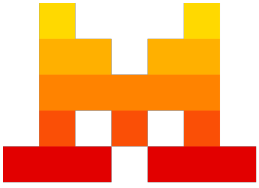
Motivation

Ecosystems

Ecosystems



Where are these logos from?



Where are these logos from?

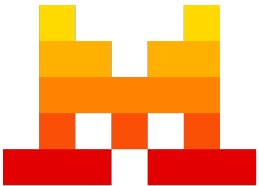


Figure 1: Mistral AIs "Le Chat"

Figure 2: OpenAIs ChatGPT

Research Gaps

- Missing quantification of AI threat
- No general usable threat model
- No threat model for offensive AI

Summary

- Complexity IT ”Ecosystems”
- Limited resources
- AI prevalence
- Duality
- Research Gaps

Summary

- Complexity IT ”Ecosystems”
- Limited resources
- AI prevalence
- Duality
- Research Gaps

Aim

Create a simplistic threat modeling method to prioritize the mitigation of offensive AI (OAI) threats.

Graph of Effort

Structure

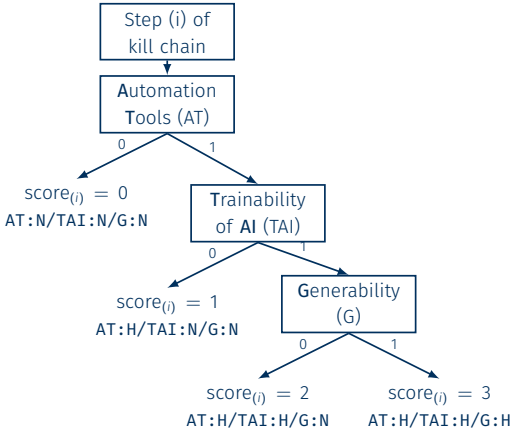


Figure 3: Graph of Effort

Killchain of Hutchins

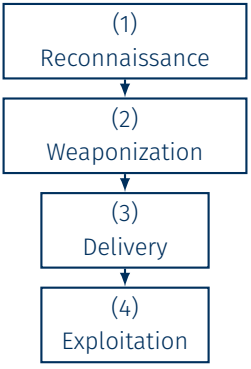
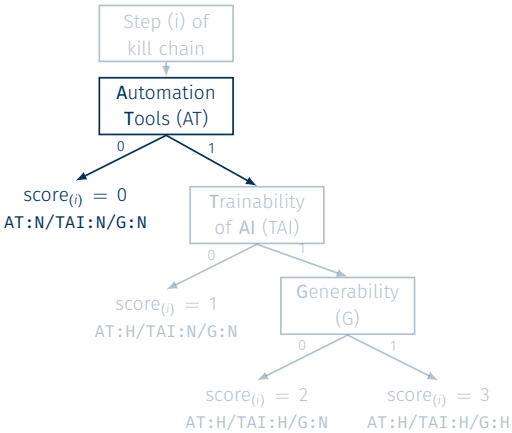


Figure 4: Steps of the intrusion kill chain according to Hutchins et. al 2011[1]

[1] E. M. Hutchins, M. J. Cloppert, R. M. Amin, et al., "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," Leading Issues in Information Warfare & Security Research, vol. 1, no. 1, p. 80, 2011

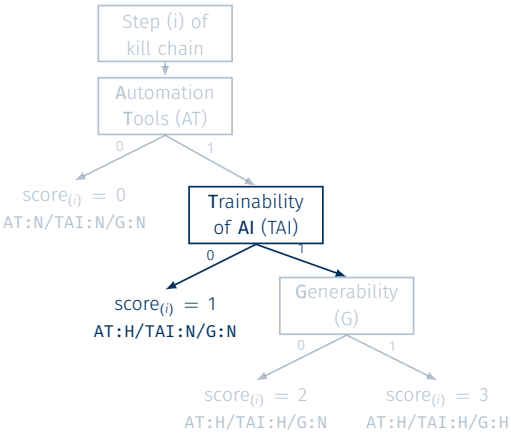
1. Automation Tools (AT)

- Do AI-based tools, AI models that are ready to use, or AI-based automatisms already exist?



2. Trainability of AI (TAI)

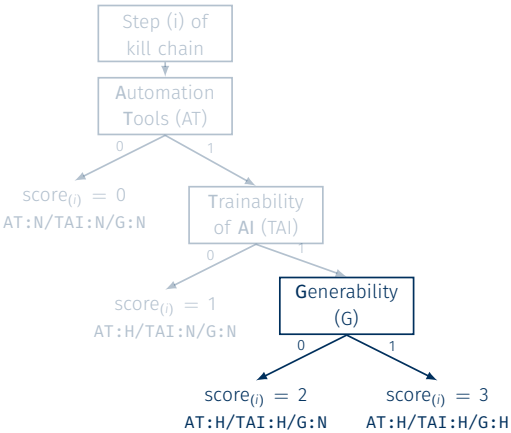
- Do ready to use datasets or even complete training setups exist which the attacker may use to generate their own AI models?



© Mehra, Prieß, Aßmuth: “Graph of Effort”

3. Generability (G)

- Are there APIs or any other tools that enable the automatic creation of data sets to create an AI model?



Score Calculation

$$\text{score}_{(i)} = \texttt{AT} + \texttt{TAI} + \texttt{G}$$

$$\text{GOE}(v) = \min_i \left\{\text{score}_{(i)}\right\}$$

Objectives

- Covering AI usage and creation aspects
- Objectivity
- Simplicity
- Flexibility
- Explainability
- usable with CVEs as part of vulnerability assessment

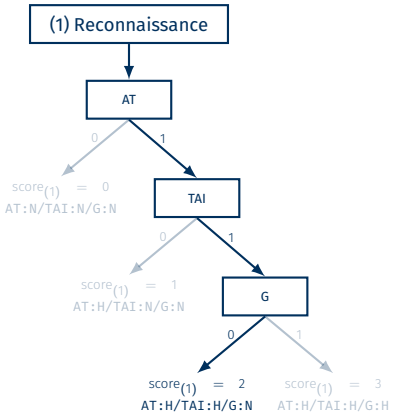
Example Scoring

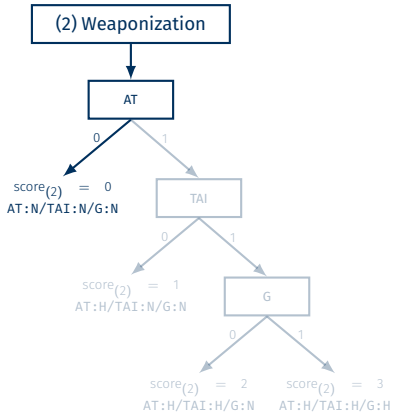
CVE-2025-1156

Description

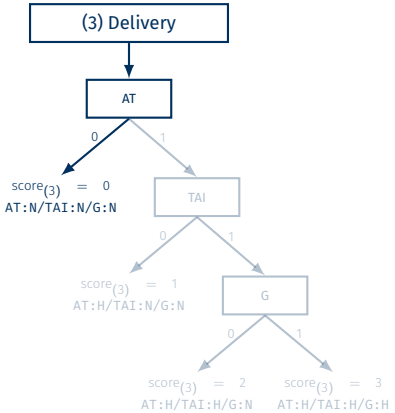
A vulnerability has been found in Pix Software Vivaz 6.0.10 and classified as critical. This vulnerability affects unknown code of the file `/servlet?act=login`. The manipulation of the argument `usuario` leads to sql injection. The attack can be initiated remotely. The exploit has been disclosed to the public and may be used. The vendor was contacted early about this disclosure but did not respond in any way.

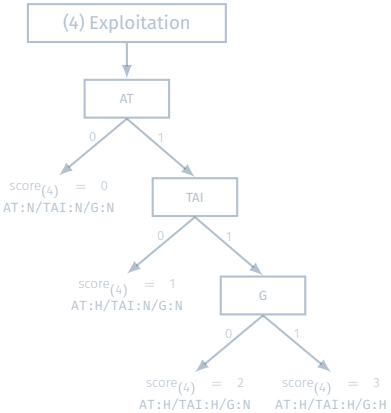
CVE-2025-1156 Scoring (1)





CVE-2025-1156 Scoring (2)





Final Score

$$\mathrm{GOE}(\text{CVE-2025-1156}) = \min\{2, 0, 0, \infty\} = 0$$

Implications

- GOE is trivial in usage
- GOE assists in quantification of AI threat (validation needed!)
- Vulnerability assessment teams need to develop AI knowledge