Andrea Vignali

AI-Empowered Cybersecurity in Cyber-Physical Systems through Natural Language Processing and Anomaly Detection

Tutor: G. Sperlì co-Tutor: S. P. Romano

Cycle: XXXVIII Year: Third

Candidate's information

- MSc degree in Computer Engineering @ DIETI – Federico II
  - Thesis: "An active learning and similarity based augmentation approach for few-shot NER applications"
- DIETI Research group/laboratory: PICUSlab and ARCLab
- PhD start date – end date: November 1, 2022 – October 31, 2025
- Scholarship type: PNRR – 352
- Partner company:
  - AKKODIS ITALY S.R.L. (former AKKA Italia s.r.l.)
  - Period in company: November 1, 2023 – May 1, 2024
- Period abroad: September 2, 2024 – February 28, 2025 @ Massachusetts Institute of Technology, Boston – CSAIL – ALFA

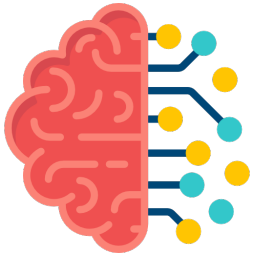
Summary of study activities

- 10 Courses
- 2 PhD schools
  - CNTC (Complex networks and telecommunications 3rd edition: Towards 6G), Como, Italy
  - Open and programmable 6G networks in the cloud/edge continuum: research challenges and experimentation tools in SLICES Research Infrastructures, Lipari, Italy
- 33 Seminars
- 3 International conferences as presenting speaker
- Tutorship activities
- Other activities:
  - Student representative for the XXXVIII cycle (ITEE)
  - PhD student representative on the Department Council
  - PhD student representative on the Joint Committee

Research field of interest

AI-Empowered Cybersecurity in Cyber-Physical Systems

- **Natural Language Processing (NLP)**
  - Data scarcity, data augmentation, few-shot learning
  - Model robustness and generalizability
  - Large Language Models (LLMs) and their evaluation



- **Anomaly Detection**
  - Multiple data sources and data fusion
  - Model robustness and generalizability
  - Use of NLP for anomaly detection in Cyber-Physical Systems (CPSs)

Research results (NLP)

- Data augmentation techniques, including reinforcement learning and active learning, have proven effective in enhancing NLP models in heavily domain-oriented task (i.e., NER)
  - These techniques improve model robustness by selecting the most effective training samples
  - LLMs stand out in general domain of application
- LLM evaluation is challenging, it varies by domain and topic, and should be both efficiency- and human-oriented, combining quantitative and qualitative metrics
- NLP techniques, when combined with anomaly detection models, have proven effective for test case prioritization

Research results (Anomaly Detection)

- In CPS, diverse data sources (e.g., network traffic, sensor and actuator readings) enable the detection of different attack types
- Different models fit different data sources
  - For the SWaT dataset, after extensive tuning and evaluation, Autoencoders perform best for network-traffic anomalies, while GANs are more effective for detecting anomalies in sensor/actuator data
- Late fusion techniques show remarkable results
  - Drawback: The overall number of false positives increases
- The use of graph neural network for sensor and actuator readings allows to capture spatio-temporal dependencies between sensors and actuators

Research products (1)

[P1]	Vincenzo Moscato, Marco Postiglione, Guido Secondulfo, Giancarlo Sperlí, Andrea Vignali, *Learning How To Augment Data: An Application To Biomedical NER,* **Knowledge Discovery in Healthcare Data (KDH)@International Joint Conferences on Artificial Intelligence (IJCAI)** Macao, China, August 19-25, 2023.
[P2]	Ilaria Bartolini, Vincenzo Moscato, Marco Postiglione, Giancarlo Sperlì, Andrea Vignali, *Data augmentation via context similarity: An application to biomedical Named Entity Recognition,* **Information Systems,** vol. 119, pp. 102291, 2023, DOI: 10.1016/j.is.2023.102291
[P3]	Ilaria Bartolini, Angelo Chianese, Vincenzo Moscato, Marco Postiglione, Giancarlo Sperlí, Andrea Vignali, *Named Entity Recognition using context similarity data augmentation,* **32nd Symposium on Advanced Database Systems (SEBD 2024),** Villasimius, Italy, June 23-26, 2024, vol. 3741, pp. 331-338, CEUR WORKSHOP PROCEEDINGS.
[P4]	Giancarlo Sperlì, Andrea Vignali, *Anomaly Detection in Cyber-Physical Systems: A Case Study on Pump Health Monitoring,* **2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW),** Jeju, South Korea, April 14-19, 2024, pp. 361-364.
[P5]	Simon Pietro Romano, Giancarlo Sperli, Andrea Vignali, *An NLP-based approach to assessing a company's maturity level in the digital era,* **Expert Systems with Applications,** vol. 252, pp. 124292, 2024, DOI: 10.1016/j.eswa.2024.124292

Research products (2)

[P6]	Vincenzo Moscato, Marco Postiglione, Giancarlo Sperli, Andrea Vignali, *ALDANER: active learning based data augmentation for named entity recognition,* **Knowledge-Based Systems,** vol. 305, pp. 112682, 2024, DOI: 10.1016/j.knosys.2024.112682.
[P7]	Roberto Canonico, Giovanni Esposito, Annalisa Navarro, Simon Pietro Romano, Giancarlo Sperli, Andrea Vignali, *An anomaly-based approach for cyber–physical threat detection using network and sensor data,* **Computer Communications,** vol. 234, pp. 108087, 2025, DOI: 10.1016/j.comcom.2025.108087.
[P8]	Roberto Canonico, Giovanni Esposito, Annalisa Navarro, Simon Pietro Romano, Giancarlo Sperlì, Andrea Vignali, *Empowered Cyber–Physical Systems security using both network and physical data,* **Computers & Security,** vol. 152, pp. 104382, 2025, DOI: 10.1016/j.cose.2025.104382.
[P9]	Roberto Canonico, Francesco Lista, Annalisa Navarro, Giancarlo Sperlí, Andrea Vignali, *Threat detection in reconfigurable Cyber-Physical Systems through Spatio-Temporal Anomaly Detection using graph attention network,* **Computers & Security,** vol. 156, pp. 104509, 2025, DOI: 10.1016/j.cose.2025.104509.
[P10]	Gabriele Dario De Siano, Anna Rita Fasolino, Giancarlo Sperlí, Andrea Vignali, *Translating code with Large Language Models and human-in-the-loop feedback,* **Information and Software Technology,** vol. 186, pp. 107785, 2025, DOI: 10.1016/j.infsof.2025.107785.

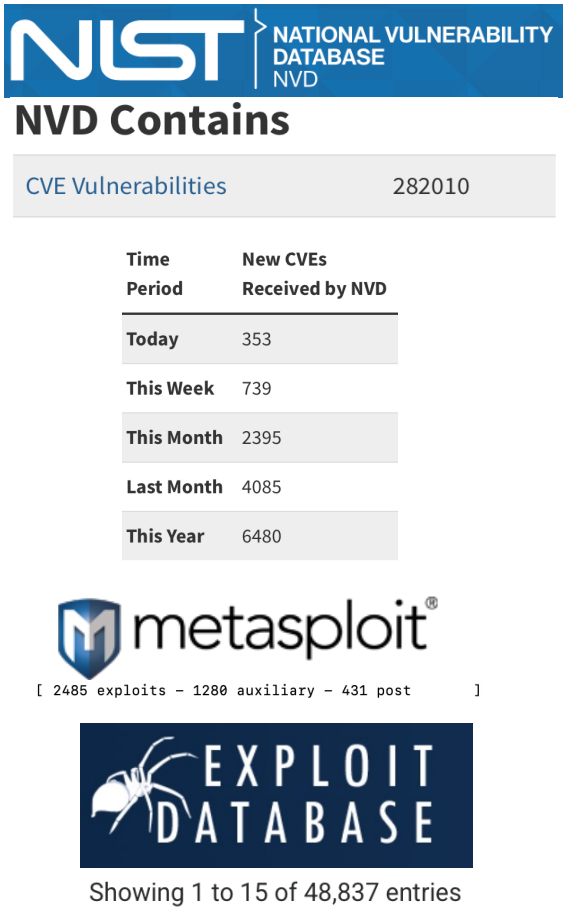
Research products (3)

[P11]	Pasquale De Falco, Giancarlo Sperlí, Marcello Vestri, Andrea Vignali, *Smart home Demand-Side Management Based on rooftop deep learning photovoltaic power forecasting,* **Sustainable Computing: Informatics and Systems,** vol. 186, pp. 101162, 2025, DOI: 10.1016/j.suscom.2025.101162.
[S1]	Miguel Tulla, Andrea Vignali, Christian Colon, Giancarlo Sperli, Simon Pietro Romano, Masataro Asai, Una-May O'Reilly, Erik Hemberg, *Hybrid Privilege Escalation and Remote Code Execution Exploit Chains,* **IEEE Transactions on Information Forensics and Security,** 2025 – Submitted. Preprint: https://doi.org/10.48550/arXiv.2504.07287
[S2]	Giovanni Officioso, Giancarlo Sperlì, Andrea Vignali, *Financial News Sentiment Meets Market Data: A Large Language Model-based Approach to Stock Price Prediction,* **Information Sciences,** 2025 – Submitted.
[S3]	Simon Pietro Romano, Giancarlo Sperlì, Mario Varlese, Andrea Vignali, *NER in the Courtroom: A Data-Driven Framework for Legal Entity Extraction,* **Information Processing & Management,** 2025 – Submitted.
[A1]	Andrea Vignali, Giancarlo Sperlì, Simon Pietro Romano, *Harnessing NLP for test case prioritization: unsupervised approaches,* **International Joint Conference on Neural Networks,** Rome, Italy, June 30 - July 5, 2025. Accepted and Presented (presenting speaker).

PhD thesis overview

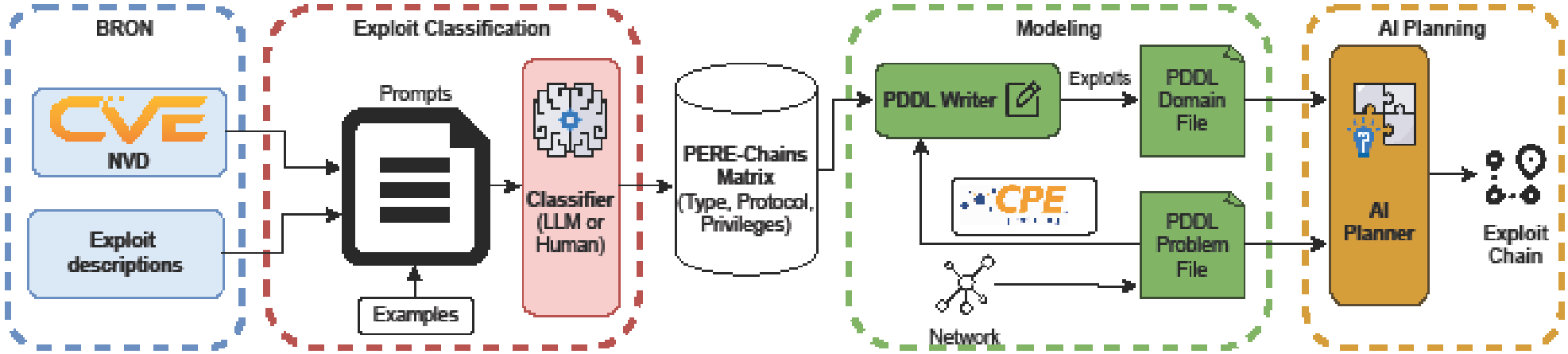
- **Problem:** Automating the discovery of exploit chains based on the structure and vulnerabilities of a given network configuration
- **Scope:** Industrial and enterprise networks
- **Objective:** Determine whether a host can be compromised with Remote Code Execution (RCE) and Privilege Escalation (PE) Exploit
- **Methodology:** Use large amounts of unstructured, unlabeled text combined with structured network descriptions to generate exploit chains

Motivation



- Compromising a network can involve chaining multiple exploits
  - Attackers have access to at least thousands of public exploits to infiltrate networks
  - The number of exploits and vulnerabilities increases daily
  - The information provided by security databases is semi-structured and hard to use as-is
- Determining exploit preconditions is non-trivial
- Real-world networks require meticulous tracking of vulnerabilities to assess whether a host can be compromised

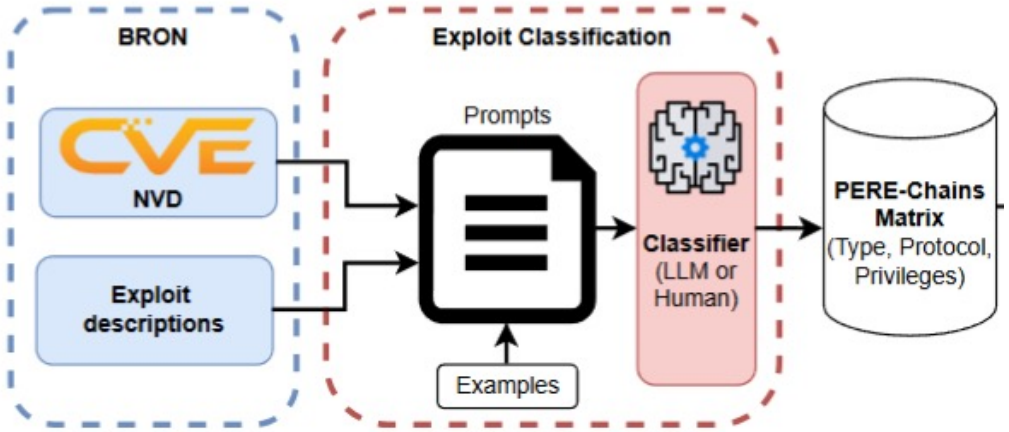
Methodology



- Exploit Classification:
  - Classifies the exploits and stores the results and the vulnerable configurations in an ALFA-Chains Matrix
- Modeling:
  - Models the problem and the domain in PDDL using the relevant vulnerable configurations and exploits
- AI Planning:
  - Derives one or more exploit chains from the problem and domain files

Exploit Classification

- We classify the exploits by:
- Type: PE, RCE
- Protocol (If Remote): TCP or UDP
- Privileges Required to run the exploit: None (N), Low (L), High (H)
- Privileges Acquired on the target: Low (L), High (H), Root (R)



- We used the privilege levels defined in the Common Vulnerability Scoring System (CVSS) Vector from v3.0 onward

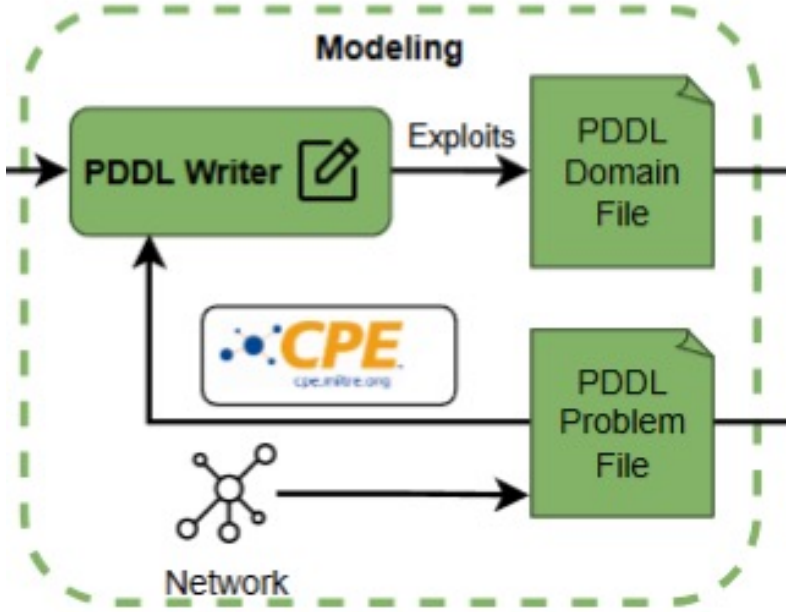
Vector: CVSS:3.0/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H

Input data:

- Exploit description text
- CVE description text
- CPE records

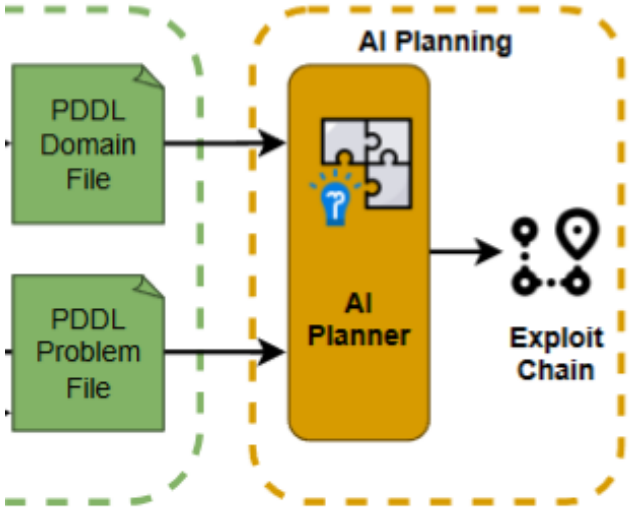
Modeling

- We defined and used a predetermined set of predicates
- Only a subset of exploits is relevant to the network
  - The PDDL Writer writes a PDDL action for each exploit that matches a vulnerable configuration in the network
- Network topology and technological stack of each host of interest are encoded in the PDDL Problem file



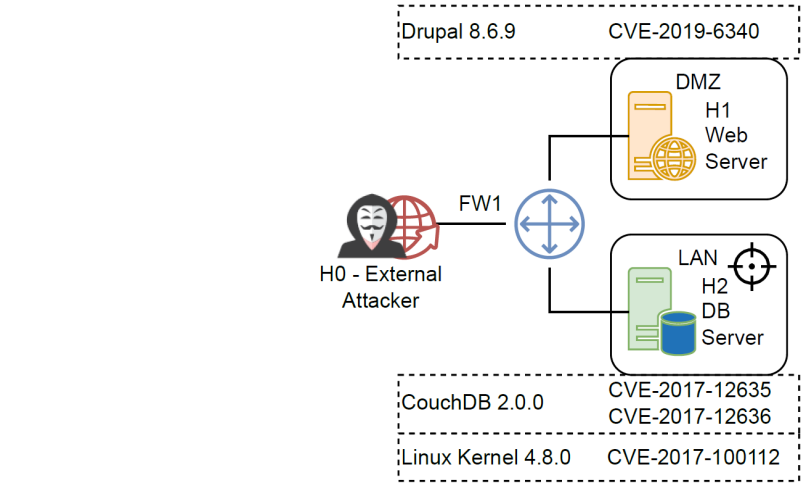
AI Planning

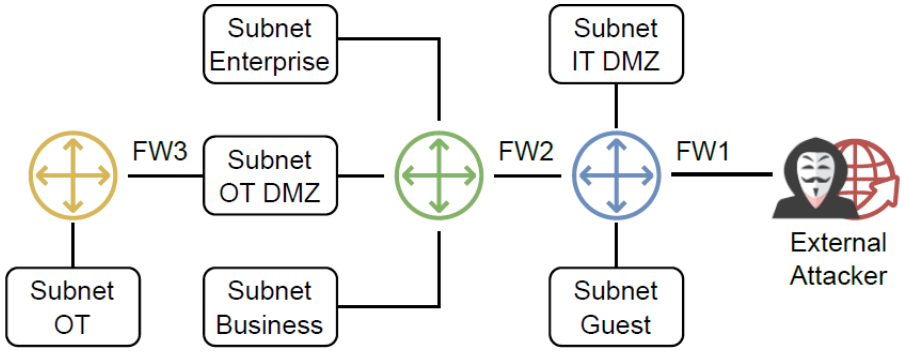
- The PDDL problem and PDDL domain file are used as input for an AI Planner
- We tested the following planners:
  - LAMA: uses a landmark heuristic and performs refinement of the solution through iterative steps
  - LAMA-first: stops at the first step of LAMA
  - K*: handles multiple optimal and suboptimal plans
  - ENHSP: handles numeric data



Experimental Networks

- Two different network architectures:
  - DMZ+LAN
  - Purdue Model
- For the Purdue Model we used 3 different hosts configurations with one vulnerability chain planted
- $Purdue_1$: 20 hosts with a total of 20 different vulnerabilities
- $Purdue_2$: 200 hosts with a total of 83 different vulnerabilities
- $Purdue_3$: 200 hosts with a total of 114 different vulnerabilities





Results (1)

- **Scaling:** We compared the performance of planners to find one exploit chain, varying the network and the number of vulnerabilities on the hosts
  - We used 1,880 exploits from Metasploit as a reference

- **Multiple plans:** In DMZ+LAN, K* found multiple exploit chains that were not previously planted
  - K* was able to find multiple exploit chains using 1,880 Metasploit exploits in all the networks proposed

Example Network				Perf.
Network	**# Hosts**	**# Actions**	**Planner**	**Duration (s)**
DMZ+LAN	2	20	LAMA-first	0.30 (0.01)
DMZ+LAN	2	20	LAMA	0.31 (0.02)
DMZ+LAN	2	20	K*	0.002 (0.000)
DMZ+LAN	2	20	ENHSP	0.07 (0.00)
Purdue_1	20	83	LAMA-first	0.43 (0.02)
Purdue_1	20	83	LAMA	1.06 (0.01)
Purdue_1	20	83	K*	0.01 (0.000)
Purdue_1	20	83	ENHSP	0.30 (0.01)
Purdue_2	200	83	LAMA-first	4.48 (0.02)
Purdue_2	200	83	LAMA	2,574 (450)
Purdue_2	200	83	K*	3.25 (0.07)
Purdue_2	200	83	ENHSP	51.61 (5.67)
Purdue_3	200	114	LAMA-first	5.77 (0.04)
Purdue_3	200	114	K*	3.16 (0.06)
Purdue_3	200	114	ENHSP	46.62 (6.47)

Network	**# Hosts**	**# Actions**	**# Plans**	**Duration (s) (Std)**
DMZ+LAN	2	20	13 (c: 3.6)	0.008 (0.000)
Purdue_1	20	83	13 (c: 6.9)	0.016 (0.001)
Purdue_2	200	83	13 (c: 6.9)	7.556 (0.127)
Purdue_3	200	114	13 (c: 6.9)	26.258 (5.846)

Results (2)

- **Generalization:** We varied the target host across DMZ+LAN and $Purdue_1$
  - Our objective is to gain ROOT privileges across the network
- Besides Metasploit, we used 1,903 Core Certified Exploit Library exploits

- **Privilege Configurations:** Applications can be configured with a specific level of privilege
  - Misconfigurations impact the number of chains that can be found
- GPT-4o showed to be a good classifier, but struggles with Privileges Acquired
  - 100 exploits from Metasploit database used as Ground Truth

Network	Host	Metasploit Plans	Core Plans
DMZ+LAN	H1	8	5
	H2	13	24
DMZ+LAN	**TOTAL**	**21**	**29**
$Purdue_1$	*data1*	30	21
	data2	30	22
	data3	30	20
	lan4	13	15
	lan5	13	15
	scada3	13	0
	scada4	13	0
	scada5	13	0
	web_server1	16	3
	web_server2	18	3
	web_server3	16	3
$Purdue_1$	**TOTAL**	**205**	**102**

Network	# Actions	LB	UB	Baseline
DMZ+LAN	20	11	13	13
$Purdue_1$	83	7	13	13
$Purdue_2$	83	7	13	13
$Purdue_3$	114	7	13	13

	Recall	Precision	F1
Exploit Type	**0.96** (0.01)	**0.97** (0.01)	**0.96** (0.01)
Protocol	0.94 (0.01)	0.95 (0.01)	0.95 (0.01)
Priv. Req.	0.92 (0.01)	0.94 (0.01)	0.93 (0.01)
Priv. Acq.	*0.75* (0.03)	*0.76* (0.02)	*0.75* (0.03)
Overall	0.69 (0.03)	0.75 (0.01)	0.71 (0.02)

Related Work

Related Work	Obes et al. (2013)	De Pasquale et al. (2024)	Ours
Planning Method	Classical planner	LLM-assisted planner	LLM-assisted planner
Planners	Metric-FF, SGPlan	PowerLifted	K*, LAMA, LAMA-first, ENHSP
Target	Networks	Single-hosts	Networks
OS	All	Unix	All
Exploit Sources	CoreImpact	GTFOBins	Core Certified Exploit Library, Metasploit
Objective achieved	Lateral movement	PE	Lateral movement + PE

Conclusions

- NLP techniques have proven essential for processing the vast amount of unstructured textual data contained in exploit databases
- The fusion of natural language data with structured network descriptions enables AI Planning to effectively and efficiently find exploit chains
- Our methodology proved scalable with respect to the number of vulnerabilities, hosts, and connections, and robust to privilege misclassification

Thanks for your attention