Karlsruhe Institute of Technology

KASTEL

Web Application Penetration Testing with Artificial Intelligence: A Systematic Review

Gustavo Sánchez*, Olakunle Olayinka†, Aryan Pasikhani†

*Karlsruhe Institute of Technology (KIT), †The University of Sheffield

The 22nd International Symposium on Network Computing and Applications (NCA 2024)

Web Application Penetration Testing with AI

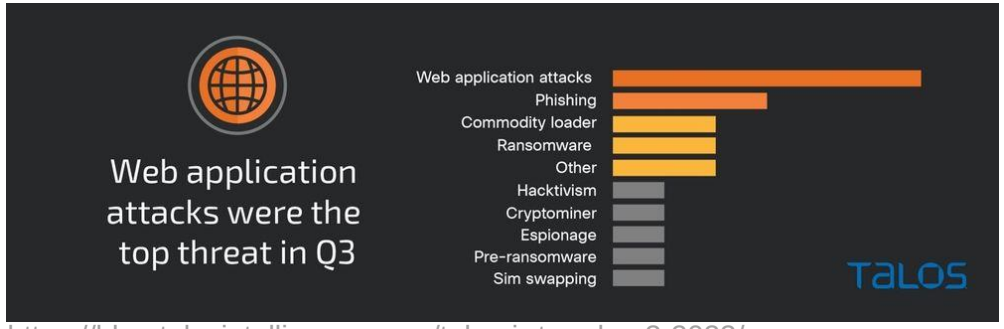
- Main Contributions
  - Identifying and examining the state of the art in this area
  - Discussing prevailing trends and challenges
  - Predicting future research directions

- Secondary Contributions
  - Address the scarcity of recent literature analyses
  - To the best of our knowledge, we are the firsts to include papers from incipient research directions (e.g., LLMs, Adversarial Attacks)

AI: Artificial Intelligence
LLMs: Large Language Models

Motivation

- Web applications are a target

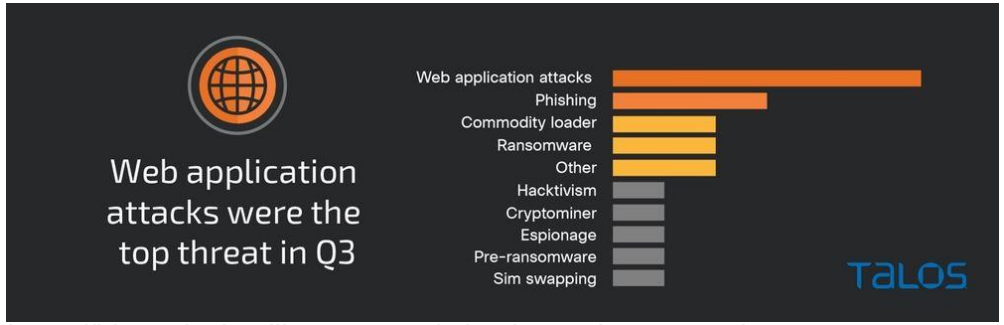


https://blog.talosintelligence.com/talos-ir-trends-q3-2023/

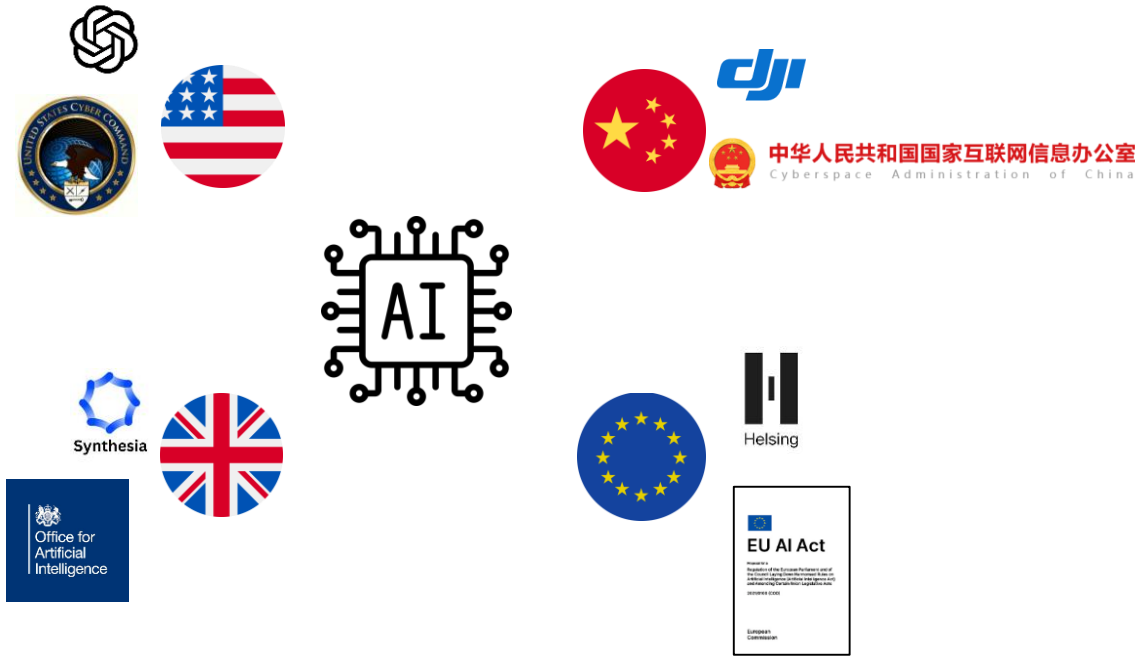
AI: Artificial Intelligence

Motivation

- Web applications are a target
- Artificial Intelligence new trends



https://blog.talosintelligence.com/talos-ir-trends-q3-2023/



AI: Artificial Intelligence

Types of Penetration Tests and Related Work

- *Threat model*: White-box vs Grey-box vs Black-box
- *Focus*: Web App vs Software pentesting
- *Objective*: Pentesting/Vulnerability Detection vs Vulnerability Prediction
- *Technique*: Static vs Dynamic

Authors	Year	Focus	Papers	Period covered
Bassi and Singh [1]	2023	Software Vulnerability Prediction	77	2007-2022
Saber *et al.* [2]	2023	General Pentesting	Undefined	Undefined
Harzevili *et al.* [3]	2023	Software Vulnerability Prediction	67	2011-2022
McKinnel *et al.* [4]	2019	General Pentesting	31	2002-2017
Our Survey	**2024**	**Pentesting Web Apps**	**49**	**2013-2024**

[1] Bassi and Singh: A systematic literature review on software vulner-ability prediction models. IEEE Access (2023)
[2] Saber, et al.,: Automated penetration testing, a systematic review. In: MIUCC. IEEE (2023)
[3] Harzevili, et al.,: A survey on automated software vulnerability detection usingmachine learning and deep learning. arXiv:2306.11673 (2023)
[4] McKinnel et al.,: A systematic literature review and meta-analysis on artificial intelligence inpenetration testing and vulnerability assessment. Computers & Electrical Engineering (2019)

Research Questions

- **RQ1:** What AI methodologies are predominantly used in web applications penetration testing, and for what specific purposes?;
- **RQ2:** How do AI-driven web application pentesting tools compare in effectiveness and efficiency to traditional methods?;
- **RQ3:** What are the recognized limitations and challenges for AI-driven web applications pentesting tools as identified in the literature?.

- Our study focuses on cybersecurity research with an offensive approach

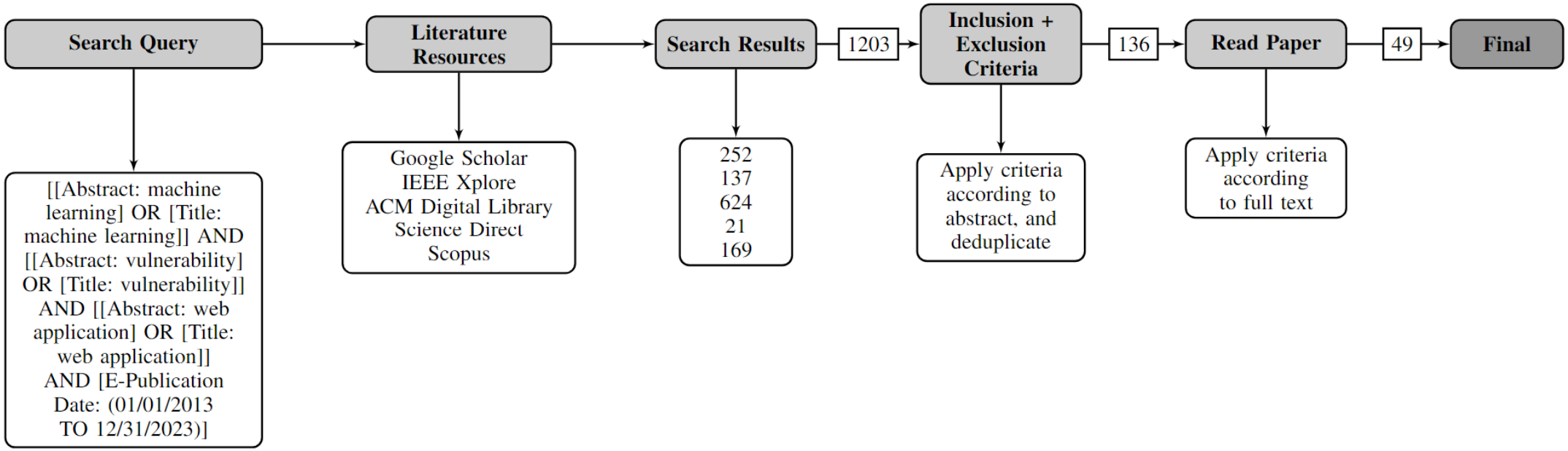
AI: Artificial Intelligence

Selection Criteria

- Inclusion Criteria
  - Employ AI methods
  - To find Web Applications vulnerabilites
  - Empirical evaluation
  - Peer-reviewed
- Exclusion criteria
  - Pentesting other domains
  - Unpublished work/Preprints
  - Non-empirical (e.g., other reviews)
  - Non-English

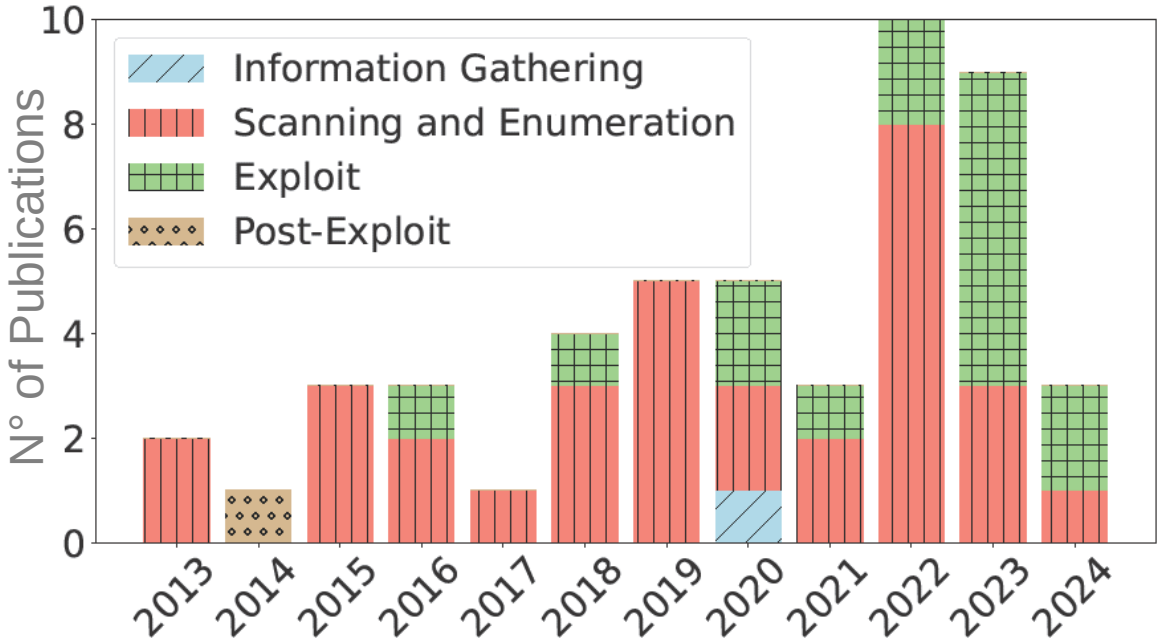
AI: Artificial Intelligence

Visual overview



Review Results I

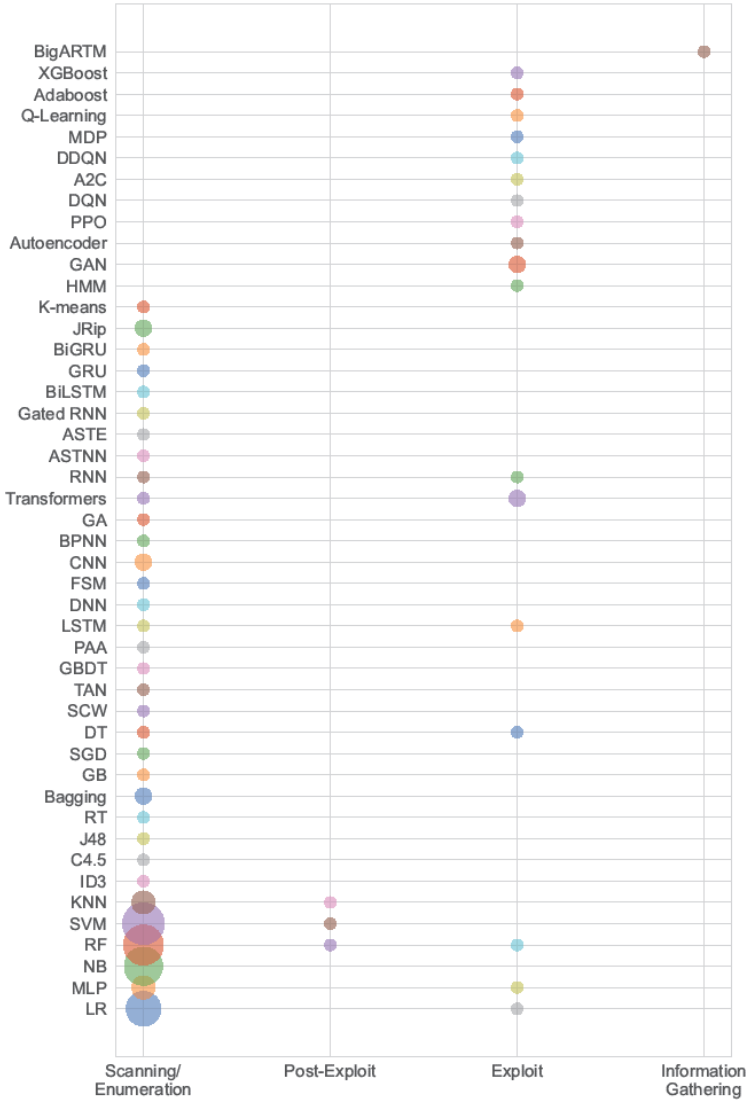
- 29 in conferences, 20 in journals
- Papers and pentesting stages investigated over the years
  - Top stage: *Scanning and enumeration*
  - More *exploit* papers as AI matures
- Dynamic analysis preferred
  - Exploitation + dynamic analysis
- Top tests: *injections*
  - *SQLi and XSS*



AI: Artificial Intelligence
SQLi: Sequel injection
XSS: Cross-site Scripting

Review Results II

- ML and NNs more frequent
  - E.g. LR, SVMs, RF, MLP
- RL least used
  - Adaptive methods still inmmature
- Recent NLP increase
  - E.g. Transformers



ML: Machine Learning
NNs: Neural Networks
LR: Logistic Regression
SVMs: Support Vector Machines
RF: Random Forest
MLP: Multi-Layer Perceptron
NLP: Natural Language Processing
RL: Reinforcement Learning

Review Results III

- Datasets
  - Supervised methods rely on annotated corpora
  - Static analysis tools leverage source code
  - Manual labeling is usually required
- Target Web Apps
  - Testing environments help to stay ethical.

Resources	URL
Various test cases	https://tinyurl.com/wh94b8t
Synthetic test cases written in PHP	https://github.com/stivalet/PHP-Vulnerability-test-suite
Archive of websites vulnerable to XSS	http://www.xssed.com/
HTTP requests from popular websites	https://github.com/alviser/mitch
Attack grammars for fuzzing	https://github.com/hongliangliang/gptfuzze
XSS payloads	https://github.com/payloadbox/xss-payload-list
Damn Vulnerable Web Application	https://github.com/digininja/DVWA

Additional Insights

- Other tools used
  - Data mining
    - E.g., Weka
  - Traditional pentesting software
    - E.g., Pixy, ZAP, Wapiti, Burp Suite
- Academic research compare solutions against commercial tools.
- Open-sourcing yields more citations

[10]*	46	2019	Scanning/Enumeration	S	SVM, DT, RF, GBDT, LR
[28]	10	2019	Scanning/Enumeration	S	RF, NB, J48
[33]	8	2019	Scanning/Enumeration	S+RL	SVM, MLP, DQN, LSTM
[17]	5	2019	Scanning/Enumeration	S	SGBT
[53]	2	2019	Scanning/Enumeration	S+U	SVM, PAA
[35]*	36	2020	Exploit	S+U	Transformers
[18]	20	2020	Scanning/Enumeration	S+U	LSTM
[70]*	5	2020	Exploit	S	GAN
[8]	4	2020	Information Gathering	S	BigARTM
[19]	2	2020	Scanning/Enumeration	S	DT, RF, MLP, NB, KNN, LR, SVM
[45]*	27	2021	Scanning/Enumeration	S	DNN
[5]*	6	2021	Exploit	U	Autoencoder
[37]	3	2021	Scanning/Enumeration	S	BPNN, GA
[48]	2	2021	Exploit	S+U	SVM, PAA, DAA
[32]*	13	2022	Exploit	RL	PPO, DQN, A2C
[31]	12	2022	Scanning/Enumeration	S	CNN
[49]	11	2022	Scanning/Enumeration	S	FSM
[43]	9	2022	Scanning/Enumeration	S	HMM
[39]	4	2022	Scanning/Enumeration	S	DT, KNN, RF, LR, SVM, LSTM, BiLSTM, GRU, BiGRU
[67]	2	2022	Scanning/Enumeration	S	LSTM, RF, GB, LR
[75]	1	2022	Scanning/Enumeration	S	Gated RNN
[47]	0	2022	Scanning/Enumeration	S	CNN, RNN, LSTM, BiLSTM
[38]	0	2022	Scanning/Enumeration	S	DT, SVM, NB, RT, RF, JRip
[36]	7	2023	Scanning/Enumeration	S	Graph CNN, RNN
[2]	2	2023	Scanning/Enumeration	S	Transformers
[68]	0	2023	Exploit	S	NB, LR, DT, RF, XGBoost
[74]	0	2023	Scanning/Enumeration	S	ASTNN, LSTM, SVM, ASTE
[73]	0	2023	Exploit	RL	DDQN
[34]*	0	2023	Exploit	U+RL	Transformers, MDP
[63]*	0	2023	Exploit	RL	Q-Learning
[26]	0	2023	Exploit	U	RF, Adaboost, SVM, RNN
[12]	0	2023	Exploit	S+RL	GAN

Answering Research Questions

- **RQ1:** What AI methodologies are predominantly used in web applications penetration testing, and for what specific purposes?
  - ML stands as the primary area of focus, complemented by NNs, NLP and RL
  - In scanning and enumeration stages: SVM and RF are popular choices for classification tasks
  - In the exploit stage, GAN and HMM are notable for their specialized applications
  - In the cases of Post-Exploit and Information Gathering: lack of focus

ML: Machine Learning
GAN: Generative Adversarial Network
NNs: Neural Networks
NLP: Natural Language Processing
SVM: Support Vector Machines
RL: Reinforcement Learnign
RQ: Research Question
HMM: Hidden Markov Chain

Answering Research Questions

- **RQ2:** How do AI-driven web application pentesting tools compare in effectiveness and efficiency to traditional methods?
  - AI-driven web application pentesting tools show promise in effectiveness and efficiency
  - Generally, the absence of standard baselines for evaluation and the diversity of approaches complicates making equitable comparisons

RQ: Research Question
AI: Artificial Intelligence
RL: Reinforcement Learning

Answering Research Questions

- **RQ2:** How do AI-driven web application pentesting tools compare in effectiveness and efficiency to traditional methods?
  - AI-driven web application pentesting tools show promise in effectiveness and efficiency
  - Generally, the absence of standard baselines for evaluation and the diversity of approaches complicates making equitable comparisons
- **RQ3:** What are the recognized limitations and challenges for AI-driven web applications pentesting tools as identified in the literature?
  - AI methods, especially supervised ML, heavily rely on high-quality annotated data
  - There is a need for common environments to evaluate new AI-based approaches
  - More open science and reproducible research needed

RQ: Research Question
AI: Artificial Intelligence
RL: Reinforcement Learning

Future Research Directions

- **Research Gaps**
  - We anticipate that future studies will focus on underrepresented OWASP vulnerabilities, such as cryptographic failures and Server-Side Request Forgery (SSRF)
- **Large Language Models (LLMs)**
  - Papers already under submission on this topic (e.g., PentestGPT)
- **Adversarial Attacks**
  - AI models can get mislead on purpose by adversaries
- **Explainability**
  - Making the decision-making processes of learning-based systems transparent and understandable to humans
- **Data Privacy**
  - Prioritise the privacy of client data, developing methods that safeguard sensitive information during and after security assessments

Conclusion

- „While AI-based tools have proven to be more efficient than traditional approaches, they still face significant challenges, such as the need for enriched data and more realistic testing environments”

Questions?

Thank you for your attention !

Contact: sanchez@kit.edu