The Role of AI in Modern Penetration Testing

J. Alexander Curtis
Department of Computer Science
Boise State University
Boise, ID. USA
alexcurtis@u.boisestate.edu

Nasir U. Eisty
Department of EECS
The University of Tennessee
Knoxville, TN, USA
neisty@utk.edu

Abstract—Penetration testing is a cornerstone of cybersecurity, traditionally driven by manual, time-intensive processes. As systems grow in complexity, there is a pressing need for more scalable and efficient testing methodologies. This systematic literature review examines how Artificial Intelligence (AI) is reshaping penetration testing, analyzing 58 peer-reviewed studies from major academic databases. Our findings reveal that while AI-assisted pentesting is still in its early stages, notable progress is underway, particularly through Reinforcement Learning (RL), which was the focus of 77% of the reviewed works. Most research centers on the discovery and exploitation phases of pentesting, where AI shows the greatest promise in automating repetitive tasks, optimizing attack strategies, and improving vulnerability identification. Real-world applications remain limited but encouraging, including the European Space Agency's PenBox and various open-source tools. These demonstrate AI's potential to streamline attack path analysis, analyze complex network topology, and reduce manual workload. However, challenges persist: current models often lack flexibility and are underdeveloped for the reconnaissance and post-exploitation phases of pentesting. Applications involving Large Language Models (LLMs) remain relatively under-researched, pointing to a promising direction for future exploration. This paper offers a critical overview of AI's current and potential role in penetration testing, providing valuable insights for researchers, practitioners, and organizations aiming to enhance security assessments through advanced automation or looking for gaps in existing research.

Index Terms—Penetration Testing; Security Testing; CyberSecurity; Artificial Intelligence;

I. INTRODUCTION

Artificial Intelligence (AI) is increasingly permeating cybersecurity, offering opportunities to enhance efficiency and effectiveness across many tasks. One domain ripe for AI-driven innovation is security penetration testing (pentesting), which has traditionally relied on manual labor and expert intuition to uncover system vulnerabilities. As system complexity and potential attack vectors continue to grow, there is an urgent need for scalable, automated solutions.

Penetration testing, commonly referred to as "pentesting," constitutes an essential methodological framework within cybersecurity risk assessment and vulnerability identification. The origins of penetration testing can be traced to the late 1960s and early 1970s, when computer security researchers began developing methodical approaches to test system vulnerabilities. It is a core cybersecurity practice in which ethical hackers simulate real-world attacks to identify security weaknesses. This security investigation procedure involves the authorization of simulated cyberattacks by experienced and trusted professionals aimed at an organization's own computer systems, networks, and applications. This is done for the purpose of assessing potential security vulnerabilities that could be exploited by malicious entities against these systems.

Pentesting enables organizations to:

- Identify exploitable vulnerabilities before malicious actors do
- Validate the effectiveness of existing defenses
- Evaluate incident response capabilities under realistic threat scenarios

Modern pentesting typically follows a structured, multistage approach. The NIST 800-115 framework outlines four key stages [1], [5], [22]:

1) **Preparation & Reconnaissance:** Collect system and network information.
2) **Discovery & Vulnerability Analysis:** Identify potential weaknesses.
3) **Exploitation:** Attempt to breach systems using discovered vulnerabilities.
4) **Reporting & Remediation:** Document findings and guide mitigation.

These stages are visualized in Fig. 1, with each stage requiring different tools and expertise. AI has the potential to enhance each step, accelerating reconnaissance, automating the selection of exploit paths, and even assisting in remediation and documentation. We aim to examine how AI contributes across these phases, particularly through the lens of the four research questions outlined in Section III.

AI offers promising capabilities for pentesting, by automating repetitive tasks, optimizing attack strategies, and uncovering novel vulnerabilities. Machine Learning (ML), a subset of AI, has already been widely deployed in cybersecurity for tasks such as anomaly detection, security testing, and code analysis. Recent progress in technologies such as RL and LLMs has extended AI's utility into earlier and more interactive phases of the software development lifecycle, positioning it as a viable force in modern pentesting.

The increasing complexity of digital ecosystems, coupled with the escalating sophistication of cyber threats, have entrenched penetration testing an indispensable component of comprehensive cybersecurity risk management strategies. As technological landscapes continue to evolve, the methodolog-
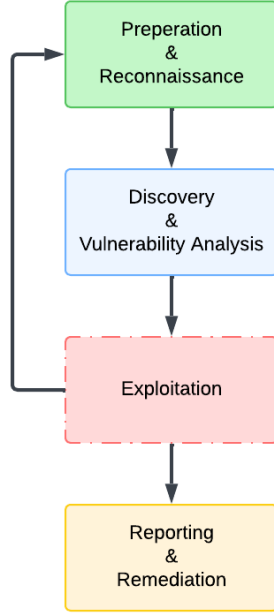
Fig. 1. Pentesting Process

ical frameworks and technological tools that support penetration testing will undoubtedly undergo continuous refinement and innovation.

Through a systematic literature review (SLR) of 58 studies, we investigate the current applications, methodologies, and benefits of AI-assisted pentesting, while identifying the challenges and limitations of current approaches. Ultimately, this research seeks to provide a comprehensive understanding of the current state of AI in pentesting and to offer insights into its future potential.

II. RELATED WORK

Several prior SLRs have addressed penetration testing, but few focus specifically on AI-assisted pentesting, and none reflect the recent surge in AI advancements.

A. Key Related Studies

The most directly relevant review is by McKinnel et al. [16], which examined AI-assisted pentesting as of 2019. However, it predates critical developments such as LLMs and primarily focuses on RL. The paper itself notes the limited scope of AI techniques explored at the time, leaving a gap in capturing today's broader AI landscape.

Ghanem and Chen [5] offer a focused review on using RL to automate pentest tasks and improve coverage. While important, their scope is narrower, rather than offering a broad synthesis of AI methodologies. It is a 2019 publication that also predates recent AI advancements.

Parveen et al. [20] published a recent SLR on pentesting methods in 2023, categorizing various tools and approaches (e.g., mobile, web, client-side), but omitting the role of AI.

Additional SLRs examine penetration testing within specific domains, such as Blockchain [12], [31], Content Management Systems [10], Industrial Control Systems [17], [23], Docker [18], and tooling [21]—but only mention AI-assisted pentesting as a future direction, not a central focus.

B. Identified Gaps and Motivation for this SLR

In light of these limitations, our paper aims to fill the gap by presenting a current, comprehensive review of AI-assisted penetration testing. Our goal is to evaluate how modern AI techniques, including RL, LLMs, Deep Learning, and other novel methodologies, are being used to support the penetration testing process in order to identify areas where future work should be focused.

III. RESEARCH METHODOLOGY

The methodology used for this research is an SLR for the current state of Artificial Intelligence technology to enhance penetration security testing. The systematic methodology of this research follows the Kitchenham and Charters [13] methodology outline to carry out an effective SLR in software engineering. We follow guidelines for inclusion, exclusion, and a compilation process based on existing literature. Our process is outlined in Fig. 2 and discussed in Section III-B.

A. Research Questions

We seek to answer the following research questions through the course of this research:

- **RQ1: How has AI been applied to pentesting?**
  This question aims to establish a baseline by examining current applications of AI in penetration testing to date, both in research and industry, helping to contextualize current capabilities and trends.
- **RQ2: What AI methodologies have been the primary focus of research related to penetration testing?**
  This research question seeks to identify which AI methodologies, such as reinforcement learning, deep learning, and natural language processing, have been explored most prominently in the context of penetration testing. Understanding where research efforts have been concentrated can help highlight promising directions, reveal underexplored areas, and inform future work aimed at improving the efficiency and effectiveness of vulnerability discovery and exploitation.
- **RQ3: Which phases of the penetration testing process are most likely to benefit from AI assistance?**
  Penetration testing consists of four distinct phases, as outlined in Fig. 1. This research question explores which of these phases stands to benefit most from the application of AI technologies.
- **RQ4: What are the key benefits and limitations associated with AI-driven penetration testing?**
  This research question aims to critically assess both the advantages and constraints of using AI in penetration testing. By evaluating the practical benefits, such as improved speed, scalability, or accuracy, alongside limitations like
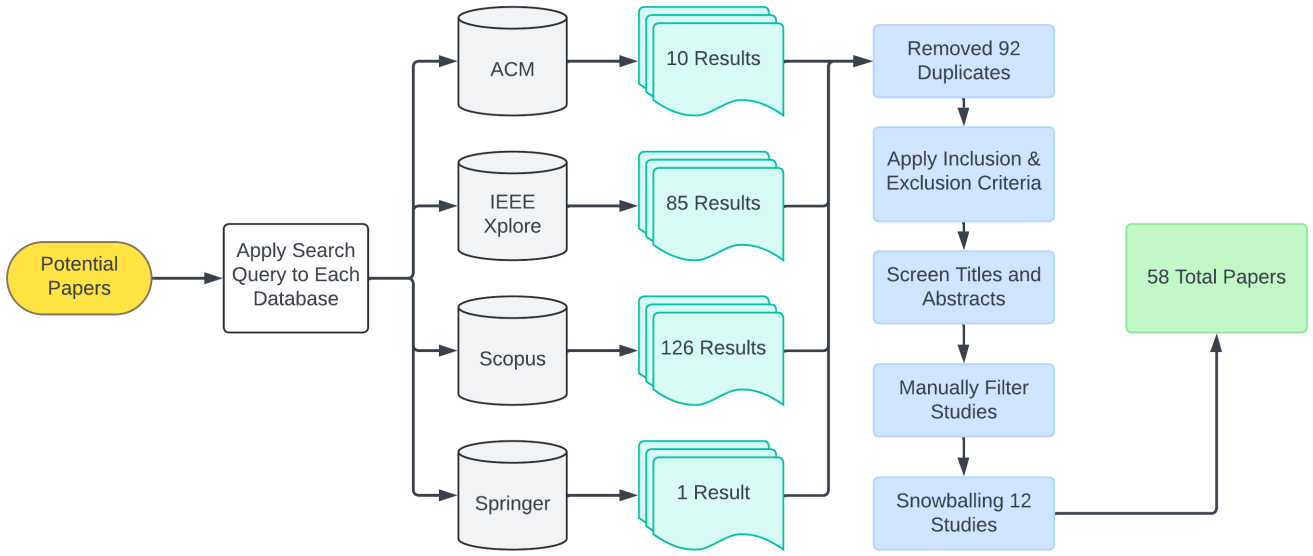
2

Fig. 2. Process Diagram for Paper Selection

interpretability, false positives, or ethical concerns, it helps establish a balanced understanding of AI's value and feasibility in this domain.

Database	Query	Results
ACM	(Abstract:("penetration testing" OR "pentest") AND Abstract:(ai OR "Artificial Intelligence" OR "Machine Learning" OR "Reinforcement Learning"))	9
IEEE	((("11Abstract":pentest OR "penetration testing") AND ("Abstract":"AI" OR "GPT" OR "LLM" OR "ML" OR "Artificial Intelligence" OR "Machine Learning" OR "Reinforcement Learning" OR "RL"))	126
Scopus	TITLE-ABS-KEY ((("pentest" OR "penetration test" OR "security test") AND AI) AND PUBYEAR > 2020 AND PUBYEAR < 2025 AND (LIMIT-TO (SUBJAREA, "COMP")) AND (LIMIT-TO (LANGUAGE, "English")))	67
Springer Link	AI Penetration Testing (Conference Paper, Article, Research article) Subdiscipline: Software engineering/programming and operating systems	19

TABLE I
SEARCH QUERIES USED FOR EACH DATABASE

B. Paper Selection

1) Search Strategy: The inclusion of literature for the review started with searching across four major academic research databases: ACM Digital Library, IEEE Xplore, Scopus, and Springer.

2) Search Criteria: The search queries that were used for each database are shown in Table I. While syntax is modified to accomodate each database, the pattern is the same,

[0]Keywords and abbreviations in search queries that yielded no additional results were removed from the table for brevity, maintaining reproducibility

requiring the mention of "Penetration Testing" or "Pentest" in the abstract AND the inclusion of some form of AI word or abbreviation including "Reinforcement Learning", "Artificial Intelligence", "Machine Learning", "Large Language Model" and their appropriate abbreviations. This approach ensured that each retrieved paper explicitly addressed both the application of AI and its relevance to penetration testing.

3) Inclusion and Exclusion Criteria: After all paper results were compiled, the 82 duplicates were removed, and then we applied inclusion and exclusion criteria to determine which papers were applicable to this review. The inclusion criteria are described in Table II, and the exclusion criteria are described in Table III.

No.	Inclusion Criteria
IC1	Full access to the document
IC2	Must be written in English
IC3	Must be published in a peer-reviewed journal or conference
IC4	Must discuss pentesting for purposes of system or network security
IC5	Must discuss AI involvement (positive or negative) in the process
IC6	The publication must answer at least one research question

TABLE II
INCLUSION CRITERIA

No.	Exclusion Criteria
EC1	Duplicate studies
EC2	Penetration Testing in fields outside computer security
EC3	Publication does not significantly contribute to the area of study

TABLE III
EXCLUSION CRITERIA

4) Data Analysis: The final study includes a total of 58 selected papers that matched the criteria and were considered

appropriate for this review.

IV. Comparison & Results

This section of the paper compares and discusses the findings we discovered based on the current literature available from our paper selection process. We address the research questions RQ1 - RQ4 below to guide and focus the comparison analysis.

A. *RQ1: Present Use of AI in Pentesting*

The findings indicate that current practical applications of AI-assisted pentesting remain limited; most uses of AI in this domain are still in the research or proof-of-concept stage. This underscores that AI-assisted penetration testing is still in its infancy, offering significant opportunities for future development and validation.

One notable real-world implementation is by the European Space Agency which developed an AI-driven pentesting platform called *PenBox* [5], [8]. This tool is tailored to detect vulnerabilities early in the development lifecycle [8]. The tool is limited specifically to space systems, and is optimized for attack patterns unique to that domain. Although its scope is narrow, PenBox demonstrates the considerable promise of AI-assisted penetration testing in operational environments with tangible benefits of cost savings and increased speed of development.

Beyond this single real-world example, several open-source and academic tools highlight how AI is being experimentally applied:

1) **Shennina-based Framework:** Karagiannis et al. [11] developed a simulation and validation tool for automated testing built on the Shennina platform.
2) **Link:** Lee et al. [14] proposed a reinforcement learning-based tool to dynamically detect XSS vulnerabilities.
3) **Pentraformer:** Wang et al. [29] introduced a reinforcement learning system for dynamic network discovery that emulates human attacker behavior.
4) **SetTron:** Yang et al. [30] presented a deep reinforcement learning model to compute efficient penetration paths without prior knowledge of the network topology.
5) **ASAP:** Chowdhary et al. [4] developed a deep neural network model to derive optimal attack policies over large enterprise networks.
6) **DUSC-DQN:** Wang et al. [27] proposed a reinforcement learning approach incorporating a Greedy-UCB algorithm to improve exploration and outperform human attackers in simulated tests.

A particularly novel application is *PenHeal*, developed by Huang and Zhu [9]. Unlike other tools focused on attack or exploration phases, PenHeal employs a two-stage pipeline using LLMs: it first identifies vulnerabilities and then guides system administrators through remediation steps. This is the only tool identified in this SLR which explicitly addresses the final phase of the penetration testing process—*Reporting & Remediation*.
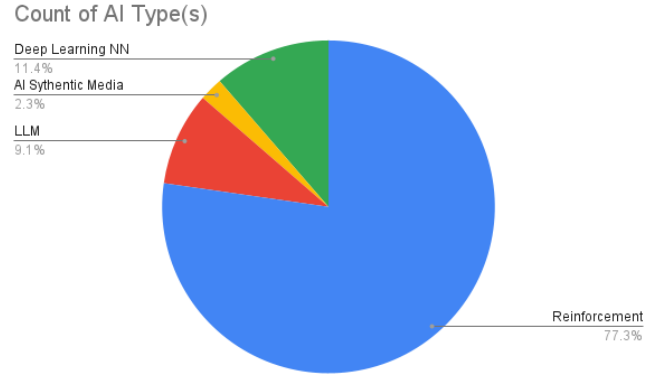

Fig. 3. Percentage of Papers Proposing Using Different Types of AI

B. *RQ2: Most Commonly Used AI Methodologies*

Artificial Intelligence encompasses a wide array of methodologies, including machine learning, reinforcement learning, deep learning, and LLMs, among others. The purpose of this research question is to identify which of these methodologies have been most prominently applied to pentesting in existing literature. This can help guide future research by highlighting where current efforts are concentrated and where opportunities remain for further exploration.

Our analysis reveals a clear dominance of Reinforcement Learning as the AI methodology with the most active research and testing. Of the 54 papers that could be classified by the penetration testing phase[1], 42 (approximately 77%) used reinforcement learning. As shown in Fig. 3, reinforcement learning accounts for the vast majority of proposed AI applications in pentesting.

A common application of reinforcement learning across these papers is the identification of optimal attack paths in large or complex network environments. These approaches often model an attacker navigating an unknown environment with minimal prior information, simulating intelligent decision-making over time.

Only four papers use LLMs as their AI technology [6], [8], [9], [19]. While LLM-based approaches are fewer, they often target different phases of the penetration testing process, such as social engineering, guidance, or remediation.

One notable distinction between RL-based and LLM-based tools lies in their deployment models. Most reinforcement learning tools are designed to run locally, making them attractive for both legitimate testers and malicious actors. In contrast, many LLM-based tools rely on access to public APIs from providers like OpenAI or Anthropic. An exception is the system developed by Gregory and Liao, which uses a locally hosted LLM based on Mistral-7B enhanced with retrieval-augmented generation (RAG) techniques [6].

---

[1]Four papers were excluded from this categorization due to their broad or conceptual focus. For example, Wang et al. [28] discuss the ethical and legal implications of AI in penetration testing rather than any specific technical implementation.
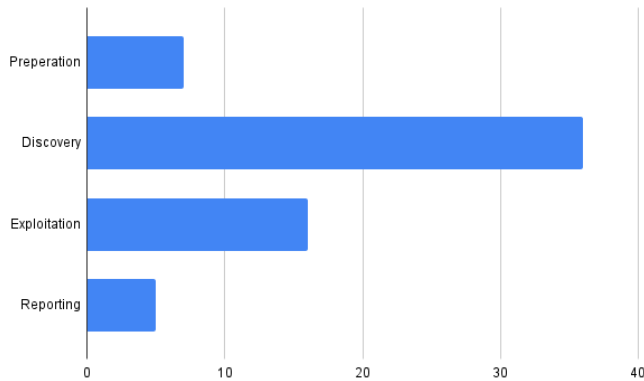
Fig. 4. Distribution of Papers Across Phases of the Penetration Testing Process

Finally, one paper stands out for its unique use of AI synthetic media (AI-SM). Soares et al. [24] introduced a method to generate realistic fake identities and documents to improve the social engineering stages of penetration testing. This was the only instance of synthetic media usage among the reviewed studies.

C. RQ3: Improvements to Process

This research question aims to identify which phases of the penetration testing process are most likely to benefit from AI assistance. As described in Section I and illustrated in Fig. 1, the process is divided into four primary phases:

1) Preparation (and Reconnaissance)
2) Discovery (and Vulnerability Analysis)
3) Exploitation
4) Reporting (and Remediation)

These phases represent broad groupings of tasks within the penetration testing lifecycle. For this literature review, each paper was analyzed and tagged based on the phase(s) it addressed. If a paper contributed to more than one phase, it was tagged accordingly. While some studies covered the full process, the majority focused on a single phase.

The distribution of research contributions across these phases is shown in Fig. 4. The results clearly show that the majority of AI-focused research targets the *Discovery* phase, followed by the *Exploitation* phase. In contrast, significantly fewer studies address the *Preparation* or *Reporting* phases, with only seven and five papers respectively. This suggests a potential gap in research and highlights areas where AI assistance could be further explored and applied.

It is worth noting that the dominance of the Discovery phase is partially driven by a concentration of studies using reinforcement learning to optimize attack paths in network environments. This has created a dense cluster of research in a single subdomain. However, this does not imply that all areas of the Discovery phase are saturated; many subtopics, such as vulnerability identification beyond network topology, remain underexplored and present opportunities for future work.

D. RQ4: Benefits and Limitations for Human/AI Assistance

This research question aims to evaluate the benefits and limitations of integrating AI into the penetration testing process. Key potential benefits include increased efficiency, improved accuracy, discovery of novel vulnerabilities, and enhanced accessibility for less experienced testers.

The most frequently cited benefit across the reviewed literature is AI's ability to accelerate time-consuming aspects of penetration testing [6], [21]. Traditional pentesting is a laborious process that requires engineers to attempt numerous potential attack vectors, most of which fail, before identifying a viable exploit. As modern infrastructure becomes more complex and distributed, the time and expertise required to perform effective tests have grown. Furthermore, network architectures and server systems are more advanced, and there are more resources in the mix that require exploration (load balancers, CDNs, network switches, routers, etc.) [25]. AI-Assisted Pentesting does not attempt to remove human engineers from the process as much as support them by pointing them to potential flaws, saving many hours of attempting vulnerabilities that are dead ends, or identifying attack vectors previously consider infeasible.

For example, Van Hoang et al. [25] developed a reinforcement learning system that observes repetitive tasks performed by human testers and automates them in future tests, thereby reducing repetitive workload. Similarly, Li et al. [15] introduced an AI agent that suggests optimal attack strategies based on ongoing interactions with human testers, continually refining its recommendations through feedback. Bianou and Batogna [3] created an LLM-based system called *PENTEST-AI*, designed to help novice engineers operate with the effectiveness of seasoned professionals by providing contextual guidance throughout the pentesting process. These examples highlight the broader potential for AI to democratize pentesting expertise and expand workforce capability.

AI has also been used to create realistic simulation environments for training purposes [7], [11], [26]. These environments can mimic enterprise network structures and provide safe, repeatable scenarios for engineers to develop and test their skills, ultimately helping to scale pentesting proficiency across teams.

However, the review also uncovered key limitations. One major challenge is the inflexibility of many early AI models, particularly those based on reinforcement learning. These models are often narrowly scoped and must be completely re-trained when new attack types or zero-day vulnerabilities emerge [21], limiting their adaptability in fast-evolving environments.

Another concern is the brute-force nature of many AI scanning techniques. AlMajali et al. [2] noted that many reinforcement learning-based systems rely on exhaustive exploration, which can generate excessive traffic and alert detection systems. In contrast, human testers tend to use more discreet, strategic probing.

In summary, AI holds great promise for enhancing the speed, scope, and accessibility of penetration testing. How-

ever, practical challenges, particularly model adaptability and operational subtlety, must be addressed before widespread deployment can be achieved.

V. FUTURE WORK

While this review provides a comprehensive snapshot of current AI-assisted pentesting research, the field remains nascent with significant opportunities for advancement. Two primary directions for future work are outlined below.

1) Expanding Coverage Across Pentesting Stages: Most research focuses on the *Discovery* and *Exploitation* stages, leaving *Preparation* and *Reporting* underexplored. Future efforts should broaden the role of AI throughout the entire lifecycle.

- **Intelligent Reconnaissance:** Use NLP and visual analysis to enhance early-stage system profiling.
- **Automated Reporting and Remediation:** Expand tools like PenHeal [9] to generate context-aware, actionable remediation guidance.

2) Integrating Emerging AI Technologies: Emerging innovations offer new opportunities for pentesting:

- **Advanced LLMs:** Explore multimodal LLMs for advance task processing and task-specific fine-tuning.
- **AI Agents for System Interaction:** Investigate agents capable of human-like system navigation & interaction.
- **Localized Models:** Develop deployable, private AI tools to reduce dependency on external APIs.

These directions highlight the need for broader phase coverage, richer AI-human interaction, and tools that balance capability with privacy and scalability.

VI. THREATS TO VALIDITY

A. Threats to Validity

Despite efforts to conduct a rigorous and impartial review, several threats may exist:

- **Selection and Publication Bias:** Restricting to English, peer-reviewed sources, may exclude relevant non-English or unpublished work. Positive result bias in academic publishing can also skew findings.
- **Search Limitations:** Although multiple databases and keyword variants were used, relevant studies with alternative terminology may have been missed.
- **Context Dependence:** Included studies may reflect specific organizational or research settings, limiting generalizability to broader or industry-specific contexts where unpublished advancements may exist.
- **Heterogeneity and Synthesis Risk:** The diversity of pentesting practices makes standardization difficult; conclusions may oversimplify a complex and evolving landscape.

VII. DISCUSSION & CONCLUSION

This systematic literature review provides a comprehensive examination of the current state of AI-assisted penetration testing, highlighting both its emerging potential and the significant challenges that lie ahead.

A. Key Findings

The review yielded several core insights:

- **Immature Landscape:** AI-assisted penetration testing remains in its early stages. Only one real-world application was identified, PenBox from the European Space Agency [8].
- **Dominance of Reinforcement Learning:** Reinforcement Learning accounts for 77% of the AI methodologies reviewed. Although this suggests strong promise, it also points to the need for further diversity in future work.
- **Phase Imbalance:** Most research targets the *Discovery* and *Exploitation* phases, leaving *Preparation* and *Reporting* phases comparatively underexplored.

B. Technological Implications

AI offers compelling advantages for penetration testing, particularly in enhancing speed and reducing manual workload, alongside challenges:

- **Efficiency Gains:** AI can automate repetitive tasks and optimize attack strategies, helping human testers work more effectively.
- **Tooling Limitations:** AI tools remain narrow in scope, with no single solution offering broad testing capabilities.

C. Practical Significance

AI is not positioned to replace human pentesters, but to augment their capabilities:

- **Assistive Role:** AI enhances the productivity of the tester by accelerating low value tasks and generating higher value insights.
- **Complexity Management:** As modern systems become more complex, AI offers scalable strategies to maintain a good security posture.

D. Challenges and Constraints

Despite its promise, AI-assisted pentesting faces important challenges:

- **Technical Gaps:** Existing models lack generalizability across diverse systems and require retraining for new contexts.
- **Ethical Risks:** The automation of offensive security tasks raises concerns about misuse, transparency, and accountability.

E. Conclusion

AI stands at the cusp of revolutionizing penetration testing. While the current landscape is characterized by limited real-world applications and predominantly theoretical research, the potential is immense. As cyber threats continue to evolve in complexity, AI-assisted penetration testing represents a critical frontier for emerging cybersecurity research. AI can continue to enable pentesters to produce more thorough and advanced results in a more time-efficient manner, ultimately propelling this critical field forward. AI is not a replacement for penetration testers, but a tool to empower them.

VIII. Data Availability

The data underlying the findings of this study will be made accessible to the research community and can be found at https://figshare.com/account/articles/29143250?file=54795278. Currently, the data is shared privately but will be made publicly available upon acceptance of the manuscript.

References

[1] NIST SP 800-115 — nist.gov. https://www.nist.gov/privacy-framework/nist-sp-800-115, 2021. [Accessed 28-11-2024].

[2] A. AlMajali, L. Al-Abed, R. Mutleq, Z. Samamah, A. A. Shhadeh, B. J. Mohd, and K. M. Ahmad Yousef. Vulnerability exploitation using reinforcement learning. In *2023 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pages 281–286. IEEE, May 2023.

[3] S. G. Bianou and R. G. Batogna. PENTEST-AI, an LLM-powered multi-agents framework for penetration testing automation leveraging mitre attack. In *2024 IEEE International Conference on Cyber Security and Resilience (CSR)*, volume 66, pages 763–770. IEEE, Sept. 2024.

[4] A. Chowdhary, D. Huang, J. S. Mahendran, D. Romo, Y. Deng, and A. Sabur. Autonomous security analysis and penetration testing. In *2020 16th International Conference on Mobility, Sensing and Networking (MSN)*, pages 508–515. IEEE, Dec. 2020.

[5] M. C. Ghanem and T. M. Chen. Reinforcement learning for efficient network penetration testing. *Information (Basel)*, 11(1):6, Dec. 2019.

[6] J. Gregory and Q. Liao. Autonomous cyberattack with security-augmented generative artificial intelligence. In *2024 IEEE International Conference on Cyber Security and Resilience (CSR)*, pages 270–275. IEEE, Sept. 2024.

[7] W. Hao, C. Shen, X. Yang, and C. Wang. Intelligent penetration and attack simulation system based on attack chain. In *2022 15th International Symposium on Computational Intelligence and Design (ISCID)*, pages 204–207. IEEE, Dec. 2022.

[8] A. Happe and J. Cito. Getting pwn'd by AI: Penetration testing with large language models. In *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pages 2082–2086, New York, NY, USA, Nov. 2023. ACM.

[9] J. Huang and Q. Zhu. PenHeal: A two-stage LLM framework for automated pentesting and optimal remediation. In *Proceedings of the Workshop on Autonomous Cybersecurity*, pages 11–22, New York, NY, USA, Nov. 2023. ACM.

[10] R. S. Jagamogan, S. A. Ismail, N. Hafizah, and H. Hafiza Abas. A review: Penetration testing approaches on content management system (CMS). In *2021 7th International Conference on Research and Innovation in Information Systems (ICRIIS)*, pages 1–6. IEEE, Oct. 2021.

[11] S. Karagiannis, C. Fusco, L. Agathos, W. Mallouli, V. Casola, C. Ntantogian, and E. Magkos. AI-powered penetration testing using shennina: From simulation to validation. In *Proceedings of the 19th International Conference on Availability, Reliability and Security*, volume 2017, pages 1–7, New York, NY, USA, July 2024. ACM.

[12] Y. Kissoon and G. Bekaroo. Detecting vulnerabilities in smart contract within blockchain: A review and comparative analysis of key approaches. In *2022 3rd International Conference on Next Generation Computing Applications (NextComp)*, pages 1–6. IEEE, Oct. 2022.

[13] B. Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26, 2004.

[14] S. Lee, S. Wi, and S. Son. Link: Black-box detection of cross-site scripting vulnerabilities using reinforcement learning. In *Proceedings of the ACM Web Conference 2022*, New York, NY, USA, Apr. 2022. ACM.

[15] Q. Li, R. Wang, M. Zhang, F. Shi, Y. Shen, M. Hu, B. Guo, and C. Xu. An intelligent penetration testing method using human feedback. *IEEE Trans. Industr. Inform.*, 20(7):9109–9119, July 2024.

[16] D. R. McKinnel, T. Dargahi, A. Dehghantanha, and K.-K. R. Choo. A systematic literature review and meta-analysis on artificial intelligence in penetration testing and vulnerability assessment. *Comput. Electr. Eng.*, 75:175–188, May 2019.

[17] N. Mohamed, A. A. Ahmed, A. Alsharif, and H. J. ElKhozondar. Employing AI-driven drones and advanced cyber penetration tools for breakthrough criminal network surveillance. In *2023 IEEE 9th International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, pages 1–6. IEEE, Nov. 2023.

[18] D. Mubanda, N. Mandela, T. Mbinda, and C. Ayesiga. Evaluating docker container security through penetration testing: A smart computer security. In *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*, pages 415–419. IEEE, Nov. 2023.

[19] T. Naito, R. Watanabe, and T. Mitsunaga. LLM-based attack scenarios generator with IT asset management and vulnerability information. In *2023 6th International Conference on Signal Processing and Information Security (ICSPIS)*, pages 99–103. IEEE, Nov. 2023.

[20] M. Parveen and M. A. Shaik. Review on penetration testing techniques in cyber security. In *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, pages 1265–1270. IEEE, Aug. 2023.

[21] V. Saber, D. ElSayad, A. M. Bahaa-Eldin, and Z. Fayed. Automated penetration testing, a systematic review. In *2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pages 373–380. IEEE, Sept. 2023.

[22] H. M. Z. A. Shebli and B. D. Beheshti. A study on penetration testing process and tools. In *2018 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, pages 1–7. IEEE, May 2018.

[23] P. Sindhwad and F. Kazi. Exploiting control device vulnerabilities: Attacking cyber-physical water system. In *2022 32nd Conference of Open Innovations Association (FRUCT)*, pages 270–279. IEEE, Nov. 2022.

[24] N. Soares, S. Seiden, I. Baggili, and A. Webb. On the application of synthetic media to penetration testing. In *The 2nd Workshop on the security implications of Deepfakes and Cheapfakes*, pages 1–10, New York, NY, USA, July 2023. ACM.

[25] L. Van Hoang, N. X. Nhu, T. T. Nghia, N. H. Quyen, V.-H. Pham, and Phan The Duy. Leveraging deep reinforcement learning for automating penetration testing in reconnaissance and exploitation phase. In *2022 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pages 41–46. IEEE, Dec. 2022.

[26] C. Wang, C. Redino, R. Clark, A. Rahman, S. Aguinaga, S. Murli, D. Nandakumar, R. Rao, L. Huang, D. Radke, and E. Bowen. Leveraging reinforcement learning in red teaming for advanced ransomware attack simulations. In *2024 IEEE International Conference on Cyber Security and Resilience (CSR)*, pages 262–269. IEEE, Sept. 2024.

[27] P. Wang, J. Liu, X. Zhong, G. Yang, S. Zhou, and Y. Zhang. DUSC-DQN:an improved deep Q-network for intelligent penetration testing path design. In *2022 7th International Conference on Computer and Communication Systems (ICCCS)*, pages 476–480. IEEE, Apr. 2022.

[28] W.-C. Wang. Legal, policy, and compliance issues in using AI for security: Using taiwan's cybersecurity management act and penetration testing as examples. In *2024 16th International Conference on Cyber Conflict: Over the Horizon (CyCon)*, pages 161–176. IEEE, May 2024.

[29] Y. Wang, S. Liu, W. Wang, C. Zhu, C. Fan, K. Huang, and C. Chen. PentraFormer: Learning agents for automated penetration testing via sequence modeling. In *2024 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics*, pages 551–558. IEEE, Aug. 2024.

[30] Y. Yang, M. Chen, H. Fu, and X. Liu. SetTron: Towards better generalisation in penetration testing with reinforcement learning. In *GLOBECOM 2023 - 2023 IEEE Global Communications Conference*, pages 4662–4667. IEEE, Dec. 2023.

[31] S. Yulianto, E. Krisnanik, and M. S. Hartawan. Strengthening IT governance in the crypto marketplace: Leveraging penetration testing and standards alignment. In *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, pages 200–205. IEEE, Nov. 2023.