

AI-Enabled Honeypot-Based Intrusion Detection for Closed-Loop Autonomous Cyber Defense: A Literature Review

Miss. Alethea Tamanna Rangayya¹, Miss. Pooja R. Tupe²

¹M.S. Cyber Security Student, ²Professor, Department of Information Technology, University of Mumbai
Vidyanagari, Kalina, Santacruz, Mumbai, Maharashtra, India

Abstract - Traditional cyber defense systems struggle against the rapid evolution of attacks in scale and complexity. Honeypots, traditionally deployed as passive deception tools for observing attacker behavior, improve detection accuracy, pattern recognition while reducing false alarm rates when integrated with artificial intelligence techniques. This paper presents a structured literature review exploring the relationships between honeypots, intrusion detection systems, artificial intelligence, deep learning and closed-loop autonomous systems. The analysis indicates that existing systems utilize open-loop designs, have partial automation, insufficient adaptability and performance constraints. This review highlights the need for the development of a closed-loop framework that enables continuous monitoring, detection and automated response to cyber threats, thereby advancing adaptive, resilient defense systems.

Key Words: Honeypots, Intrusion Detection Systems, Artificial intelligence, Machine Learning, Cybersecurity, Closed-Loop Defense, Autonomous Cyber Defense

1. INTRODUCTION

Organizations suffer major losses if cyber threats on their critical assets and infrastructure are not mitigated in real-time [1]. Cyber defense systems and strategies must be upgraded to respond to the rise in complexity of cyber-attacks, increasing attack surface, exposure to the cloud and distributed networks [2].

Most existing systems follow an open-loop design, separating the monitoring, detection and response mechanisms among varied components. This gap causes performance overhead, limited adaptability and latency in response. For example, signature-based intrusion detection systems scan network traffic against a database of known signatures [3]. These systems are incapable of identifying unknown attacks [4].

In response, artificial intelligence can be combined with the output of these systems to provide feedback mechanisms and improve pattern recognition [5], [6]. High-value attack data from honeypots is used to enhance the model's detection accuracy [7], [8]. This literature review argues for a closed-loop autonomous honeypot framework that utilizes artificial intelligence to counter modern cyber threats. The key contribution of this paper is a structured taxonomy of AI-enabled honeypot-based intrusion detection systems and

a synthesis of research gaps toward the realization of closed-loop autonomous cyber defence architectures.

2. BACKGROUND

2.1 Honeypot Systems

A honeypot is an intentionally created system or resource that mimics legitimate targets or applications but contains no real sensitive data [7], [8]. The goal of a honeypot is to lure attackers and gather threat intelligence [7], [8] based on their tactics, techniques, and methods. Studying this data helps organizations strengthen their defenses [7], [8] against such cyber threats.

Based on the level of interaction [7], [8], honeypots are divided into two main types. Low-Interaction Honeypots simulate basic services, are easier to deploy and maintain, but offer limited intelligence about the attackers [7], [8]. High-Interaction Honeypots are fully functional systems, riskier, but engage the attacker with extensive interactions and gather a wealth of information about intrusion techniques [7], [8].

Honeypots are also classified by purpose [7], [8]. Research Honeypots are used by researchers for analysis and strategy development [7], [8]. Identifying threat patterns contributes to a broader understanding of cybersecurity threats. Production Honeypots are deployed in the organization to detect and distract attackers within an internal network [7], [8]. These honeypots not only provide insight but also protect the critical assets [7], [8].

Despite their advantages, honeypots face resource-intensive setups, limited visibility, higher cost, and complexity [7], [8].

2.2 Intrusion Detection Systems

Intrusion Detection Systems (IDS) are security mechanisms that monitor network or system traffic for malicious behavior, suspicious activities or policy violations [9], [3].

IDS uses two detection methods, namely signature-based and anomaly-based [9], [2], [3]. Signature-Based IDS compares traffic patterns to known attack signatures. The IDS is effective for known threats but cannot detect threats outside its database [9], [3]. Anomaly-Based IDS uses machine learning or statistical models to detect deviations from the normal behavior [2], [4], [6]. This method is used to identify unknown threats or zero-day exploits. Hybrid IDS

integrates multiple methods to broaden its scope and threat coverage [6], [10].

IDS is limited by its reliance on known signatures, high volume of false alerts, performance overhead in large networks and limited adaptability to evolving threats [9], [2], [3], [11].

2.3 Artificial Intelligence in Intrusion Detection Systems

Artificial Intelligence (AI), comprising Machine Learning (ML) and Deep Learning (DL), has revolutionized the capabilities of IDS [4], [5], [12], [13], [6]. These algorithms help identify and mitigate malicious activities or anomalies with better accuracy and efficiency [5], [12], [13].

AI-based IDS typically uses supervised learning with honeypot-generated labelled datasets to distinguish between normal and suspicious behavior [4], [5], [12], [13], [10]. Unsupervised learning can be employed to detect unknown attacks by modelling normal behavior [11], [13], [6].

Deep learning uses neural networks for pattern recognition. DL focuses on reducing manual feature engineering and improving real-time detection capabilities [5], [12], [13].

AI is beneficial to IDS but is quickly affected by data bias, model drifting, adversarial attacks, and black-box model concerns [4], [5], [11], [13], [6].

2.4 Closed-Loop Autonomous Cyber Defense Systems

Closed-loop autonomous system follows a continuous cycle of sensing deviations from the stated normal, analyzing the said change, perform decision-making and execute automated cyber defense responses. IDS systems that follow closed-loop design can identify and respond to evolving cyber threats dynamically [9], [2], [4], [6]. These systems improve detection accuracy by refining the models based on observed network activity [4], [6]. Feedback learning is employed based on previous outputs to improve future decisions. Attack data collected from honeypots and monitored environments is fed into these learning loops [7], [8], [14]. This helps regulate detection rules and improves the system's ability to recognize unseen attack patterns [4], [5], [6].

In practical environments, closed-loop autonomous systems are automated cybersecurity frameworks that incorporate monitoring, threat detection and real-time response to cyber threats without any human intervention [4], [6]. They perform automated mitigation actions such as isolating compromised hosts and blocking malicious network traffic to prevent lateral movement within organizations [15], [16]. These systems are deployed in enterprises, IoT ecosystems, and cloud infrastructures where human-driven responses are often too slow to counter fast-moving threats [1], [15], [16].

However, the effectiveness of these systems can be affected by adversarial manipulation, where attackers attempt to

mislead detection models or poison learning processes, leading to incorrect or unstable automated responses [17], [18].

Overall, closed-loop autonomous systems integrate intrusion detection, honeypot-derived intelligence, and AI-driven automation into a unified adaptive cybersecurity framework capable of continuous learning and real-time defense [4], [6], [19].

3. EVOLUTIONARY TAXONOMY OF HONEYPOT-BASED INTRUSION DETECTION SYSTEMS

The taxonomy presented in this section is derived from the analysis and synthesis of the reviewed literature, capturing the evolutionary progression of honeypot-based intrusion detection systems across different levels of intelligence and automation. Fig -1 illustrates the four-stage evolutionary framework of AI-enabled honeypot-based intrusion detection systems.

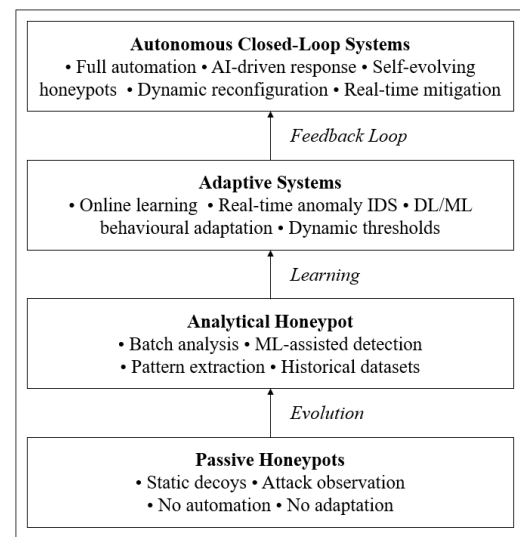


Fig -1: Evolutionary taxonomy of AI-enabled honeypot-based intrusion detection systems

3.1 Passive Honeypot Systems

Early honeypots were decoy systems designed to simulate services or assets [7], [8]. These were initially deployed by researchers to passively observe attacker behaviour [7], [8]. These non-reactive systems were low interaction honeypots [7], [8]. Passive honeypots had no adaptation to attacker behaviour. These systems were static, could not update them nor learn from any feedback [7], [8], [14].

3.2 Analytical Honeypot Systems

Analytical Honeypot is focused solely on analyzing the data collected [7], [14]. Honeypot collected data is used for offline or semi-offline analysis [8], [14]. Compared to the previous

type, these systems focus on pattern extraction and interpretation [9], [2].

Statistical analysis and ML techniques are applied to identify attack patterns, behavioural trends and anomalies in traffic/data [2], [4], [11]. These systems analyze data in batches or delayed processed mode and rely heavily on historical attack datasets [3], [10]. Basic ML-based methods such as data mining techniques are applied on the attack data [4], [6]. However, lack of real-time response loop still remains a major disadvantage [9], [3], [11].

3.3 Adaptive Honeypot Systems

Adaptive Honeypot Systems focus on continuous or online feedback learning, i.e., these systems learn from incoming traffic, not just from historical data [2], [4], [11]. Unlike offline analytical systems, these systems adjust their detection thresholds, anomaly scoring, rule sets and feature weights, based on new observations [9], [2], [4].

Machine learning and deep learning techniques are used for adaptive anomaly detection [4], [12], [13], [6]. However, these systems struggle with concept drift handling and require external supervision or retraining cycles [4], [11], [6].

3.4 Autonomous Closed-Loop Honeypot Systems

The most advanced stage in this evolution is the autonomous closed-loop honeypot systems [19], [15], [16], [17], [18]. These systems integrate detection, decision and response into a single automated loop. The output of the analysis directly influences the next system's action with minimal to no manual intervention. The system may perform automated actions such as isolating infected nodes, dynamically reconfiguring honeypots, updating firewalls or access rules and deploying decoy variations automatically.

Deep learning models are used for advanced attack classification and behaviour prediction [5], [12], [13], [6]. This learning can enhance detection accuracy, response policies and deception effectiveness.

Honeypots become self-evolving as they can mutate behaviour, change service profiles dynamically and adapt deception strategies based on attacker behaviour [20], [15], [16].

4. LITERATURE REVIEW

4.1 Foundational Honeypot and IDS Research

Spitzner (2003) [7] introduced passive honeypots as deception-based security systems designed to observe attacker behaviour and collect threat intelligence. The study demonstrated how interaction data gathered from attackers could help organizations understand intrusion patterns and improve defensive strategies.

Building on this idea, Provos (2004) [8] proposed the use of virtualization for deploying scalable virtual honeypots. By reducing hardware dependency, the framework enabled multiple honeypots to operate simultaneously in real-world environments, improving large-scale attack data collection.

Similarly, Nawrocki et al. (2016) [14] reviewed different honeypot architectures and emphasized their role in intrusion analysis and cyber threat monitoring. The study also highlighted the importance of maintaining high-quality and consistent datasets for effective security analysis.

In parallel, Liao et al. (2013) [9] examined various intrusion detection approaches, including signature-based and anomaly-based detection techniques. The research discussed how IDS frameworks monitor malicious network activities while also identifying scalability and performance limitations in large network environments.

Focusing specifically on anomaly detection, García-Teodoro et al. (2009) [2] explored statistical and behavioural methods for identifying abnormal network traffic. Although anomaly-based IDS demonstrated the ability to detect unknown attacks, the study also reported high false alarm rates caused by poor training data quality.

Earlier foundational work by Denning (1987) [3] established the core principles of modern intrusion detection systems using audit records and statistical thresholds to identify system misuse. This research laid the groundwork for later developments in intelligent and adaptive intrusion detection mechanisms.

Collectively, these studies established the foundation of honeypot systems and intrusion detection frameworks. Early defense systems primarily focused on passive monitoring, attack observation, and signature or anomaly-based detection. While these approaches improved threat visibility and attack analysis, they lacked adaptive learning and automated response capabilities. These limitations later encouraged the integration of artificial intelligence into intrusion detection systems.

4.2 Machine Learning and Deep Learning in IDS

As cyber threats became more complex, researchers increasingly explored artificial intelligence techniques to improve intrusion detection accuracy and adaptability. Buczak and Guven (2016) [4] reviewed machine learning approaches such as classification, clustering, and anomaly detection in cybersecurity applications. Their work emphasized that IDS performance heavily depends on the quality and representativeness of training datasets, making honeypot-generated attack data highly valuable for model training.

Extending this direction, Ferrag et al. (2020) [5] examined deep learning architectures including CNNs and RNNs for cyber threat detection. The study showed that deep learning models improve automated feature extraction and detection

accuracy compared to traditional machine learning methods, particularly in high-dimensional traffic environments.

However, Sommer and Paxson (2010) [11] argued that many machine learning-based IDS models perform well only under controlled research conditions. Their study highlighted the gap between experimental results and real-world deployment environments, where noisy and unpredictable traffic often affects detection reliability.

To reduce dependency on manual feature engineering, Shone et al. (2018) [12] proposed a deep learning framework combining autoencoders with neural networks for intrusion classification. Although the approach improved feature learning, it also required large training datasets and high computational resources.

Likewise, Yin et al. (2017) [13] developed an RNN-based intrusion detection model capable of capturing sequential traffic behaviour more effectively than conventional methods. This approach was particularly useful for identifying time-dependent attack patterns in dynamic network environments.

Xin et al. (2018) [6] further explored the application of machine learning and deep learning in cybersecurity and discussed their advantages in pattern recognition and anomaly detection. At the same time, the study identified challenges related to scalability, robustness, and real-time deployment.

Supporting these AI-based approaches, Moustafa and Slay (2015) [10] introduced the UNSW-NB15 dataset as a benchmark dataset for evaluating machine learning and deep learning-based intrusion detection systems. The dataset remains widely used for training and validating cybersecurity detection models.

Overall, machine learning and deep learning improve intrusion detection accuracy and feature extraction compared to traditional IDS methods. However, issues such as dataset dependency, high computational cost, concept drift, and limited real-world adaptability still restrict effective deployment, highlighting the need for more adaptive and autonomous cyber defense systems.

4.3 IoT and Adaptive Intelligent Honeypot Systems

As IoT devices became increasingly connected to enterprise and cloud environments, traditional intrusion detection approaches struggled to handle large-scale automated attacks. Addressing this issue, IoTPOT by Y. M. P. Pa et al. (2015) [20] introduced a lightweight honeypot framework designed specifically for IoT systems. The framework successfully captured real-world botnet activity and identified evolving attack behaviour targeting connected devices. This study demonstrated the growing importance of adaptive honeypot systems in modern distributed environments and highlighted the need for intelligent IDS frameworks capable of responding to continuously changing attack patterns.

4.4 Autonomous and Closed-Loop Cyber Defense Architectures

With the increasing speed and complexity of cyber-attacks, researchers began exploring autonomous cyber defense systems capable of real-time monitoring and response. K. Thakur et al. (2016) [1] discussed the limitations of conventional human-driven defense mechanisms in highly interconnected environments. The study emphasized that slow or fragmented response systems are often ineffective against rapidly spreading attacks targeting cloud platforms, industrial systems, and distributed enterprise networks.

To address these challenges, Alexander Kott et al. (2018) [19] proposed the Autonomous Intelligent Cyber-Defense Agent (AICA) reference architecture. The framework integrates monitoring, reasoning, decision-making, and automated response into a unified autonomous defense system. However, the study also identified challenges related to interoperability, trust management, and the reliability of autonomous decision-making.

Further extending this discussion, S. Vyas et al. (2026) [15] reviewed the real-world deployment of autonomous cyber defense systems and highlighted the gap between experimental research and operational implementation. Their findings identified scalability issues, integration complexity, and adversarial manipulation risks as major barriers to large-scale adoption.

Similarly, G. Sarraf and V. Pal (2026) [16] examined AI-driven threat detection and automated response mechanisms in cloud security environments. Their research demonstrated that AI-assisted defense systems improve detection speed and response efficiency compared to traditional rule-based approaches. However, challenges involving computational overhead, model explainability, and adaptive attack evasion still remain.

Focusing on explainability and trust, M. M. Ismail et al. (2025) [17] explored AI and soft-computing techniques for autonomous cyber defense systems. The study emphasized the importance of explainable AI in environments where automated security decisions directly impact critical infrastructure and sensitive operations.

More recently, the study "Secure Autonomous Cyber Defense with LLM Agents" (2026) [18] investigated the use of large language model-based agents for autonomous cybersecurity operations. The research discussed how intelligent agents can support intrusion analysis, adaptive planning, and automated response within closed-loop defense systems. At the same time, the study highlighted concerns related to hallucination risks, adversarial prompt manipulation, and governance constraints.

The studies show a shift from passive monitoring toward more adaptive and autonomous cyber defense systems. While recent approaches integrate AI-driven detection and automated response in closed-loop frameworks, challenges

such as explainability, scalability, adversarial robustness, and real-world deployment still limit full autonomy.

5. COMPARATIVE ANALYSIS

Existing honeypot-based intrusion detection systems differ significantly in terms of intelligence, automation, adaptability, and response capability. Early honeypot frameworks primarily focused on passive attack observation and deception-based monitoring without incorporating learning or automated mitigation mechanisms [7], [8], [14]. These works established foundational honeypot architectures and intrusion detection principles but lacked adaptive intelligence and real-time response capabilities [3], [9]. Later approaches introduced machine learning and deep learning techniques to improve intrusion detection accuracy and anomaly identification [4] - [6], [10] - [13]. However, most of these systems still operate using offline datasets and partially automated workflows, limiting adaptability in dynamic environments and real-world deployment scenarios [10], [11], [13].

More recent research explores autonomous cyber defense architectures capable of integrating monitoring, detection, reasoning, and response within feedback-driven environments [15], [16], [19]. Although these approaches demonstrate progress toward closed-loop defense, challenges related to explainability, adversarial robustness, scalability, and real-time deployment remain unresolved [17], [18], [19], [20].

Table 1 presents a comparative evaluation of representative approaches discussed in the reviewed literature.

Table -1: Comparative Analysis of Honeypot-Based IDS Approaches

Studies	AI Integration	Real-Time Capability	Feedback Learning	Automation Level	Key Challenge
Spitzner [7], Provos [8], Nawrocki et al. [14]	None	No	None	Manual	Static and non-adaptive systems
Denning [3], Liao et al. [9], García-Teodoro et al. [2]	Statistical / Rule-Based	Partial	Limited	Detection only	High false positive rates
Buczak and Guven [4], Sommer and Paxson	Machine Learning	Partial	Offline retraining	Semi-automated	Dataset dependency and concept drift

[11], Xin et al. [6], Moustafa and Slay [10]					
Ferrag et al. [5], Shone et al. [12], Yin et al. [13]	Deep Learning	Near real-time	Partial adaptation	Semi-automated	High computational overhead
Y. Pa et al. [20]	Limited AI support	No	None	Monitoring only	Lack of IDS-response integration
Kott et al. [19], Vyas et al. [15], Sarraf and Pal [16], Ismail et al. [17], Alqahtani and Ahuja [18]	Advanced AI / Autonomous Reasoning	Yes	Continuous feedback	Semi to fully autonomous	Explainability, scalability, and adversarial robustness

6. RESEARCH GAP

Across the reviewed literature, a clear gap exists in the unified integration of honeypots, AI-based intrusion detection, and automated response within a single closed-loop framework. Existing studies treat honeypots primarily as passive or semi-active data collection tools, with their outputs largely processed offline or used in separate analytical pipelines. Although machine learning, deep learning, and autonomous cyber defense models improve detection accuracy, they are generally not directly connected to real-time honeypot intelligence. This results in fragmented architectures where sensing, detection, and response operate independently, limiting real-time adaptability.

Recent work on autonomous and AI-driven cyber defense introduces partial feedback mechanisms, but lacks stable end-to-end implementation under dynamic and adversarial conditions. Current approaches do not adequately address continuous learning from live honeypot interactions while maintaining system stability and resistance to manipulation. The integration of real-time honeypot data into adaptive intrusion detection and automated response remains insufficiently explored, highlighting the need for a fully integrated closed-loop honeypot-based intrusion detection framework.

7. FUTURE RESEARCH DIRECTIONS

Future research can focus on designing real-time honeypot signal transformation pipelines that convert raw attacker interactions into structured, high-value security features such as behaviour sequences, protocol anomalies, and session-level attack intent. These features can then be directly used by lightweight intrusion detection models for faster and more accurate live classification. The main challenge is reducing processing delay while still preserving enough contextual information from attacks to support reliable detection.

Another important area is developing continuously adaptive detection models that learn directly from evolving honeypot interactions. Instead of relying on full retraining, the system should gradually update itself as new attack patterns appear in real time. This requires carefully controlled incremental learning techniques that allow the model to adapt to new threats without losing previously learned behaviour or becoming unstable during frequent updates.

A third direction is ensuring reliable learning from adversarial or misleading honeypot traffic. Since attackers actively interact with honeypots, they may generate noisy or deceptive patterns that can distort learning. Future work should focus on identifying and filtering such unreliable interactions using trust or reliability scoring methods, ensuring that only meaningful attack behaviour contributes to model updates and decision-making.

For implementation, a practical approach is a stream-based modular architecture where honeypots continuously generate live interaction data, which is processed into features, passed to an online intrusion detection model, and then connected to an automated response system that performs real-time defensive actions. This enables continuous monitoring, detection, and response within a single operational loop, as shown in Fig -2.

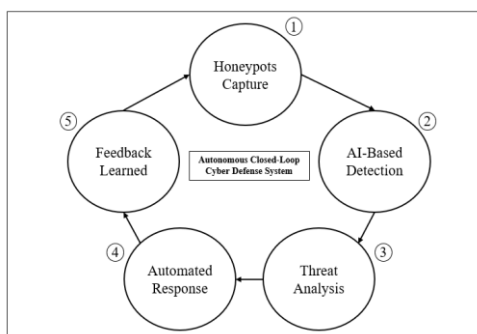


Fig -2: Closed-loop AI-enabled honeypot intrusion detection architecture with feedback-driven learning

8. CONCLUSION

The review across honeypot systems, IDS architectures, and AI-driven cyber defense models shows a clear progression from passive monitoring to partially intelligent and autonomous approaches. Honeypots consistently provide rich attacker interaction data, while machine learning and deep learning improve detection accuracy and anomaly

recognition. Despite these advancements, most systems still operate in disconnected stages where data collection, detection, and response are not tightly integrated.

Existing approaches largely depend on offline analysis, static datasets, or periodic retraining, which limits responsiveness to fast-evolving attack behaviour. Even autonomous cyber defense frameworks reported in recent literature show only partial feedback integration and lack stable coordination between detection outputs and automated mitigation actions under real-time conditions.

The overall outcome of this study points to the need for a tightly coupled, real-time honeypot-driven intrusion detection framework where continuous learning and automated response operate within a single coordinated loop.

REFERENCES

- 1) K. Thakur, M. L. Ali, N. Jiang, and M. Qiu, "Impact of Cyber-Attacks on Critical Infrastructure," in Proc. IEEE 2nd Int. Conf. Big Data Security Cloud (BigDataSecurity), New York, NY, USA, 2016, pp. 183–186.
- 2) P. García-Teodoro et al., "Anomaly-Based Network Intrusion Detection," *Comput. Secur.*, vol. 28, no. 1, pp. 18–28, 2009.
- 3) D. E. Denning, "An Intrusion-Detection Model," *IEEE Trans. Softw. Eng.*, vol. SE-13, no. 2, pp. 222–232, 1987.
- 4) L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2016.
- 5) M. A. Ferrag et al., "Deep Learning for Cyber Security Intrusion Detection," *IEEE Access*, vol. 8, pp. 165140–165175, 2020.
- 6) Y. Xin et al., "Machine Learning and Deep Learning Methods for Cybersecurity," *IEEE Access*, vol. 6, pp. 35365–35381, 2018.
- 7) L. Spitzner, *Honeypots: Tracking Hackers*. Boston, MA, USA: Addison-Wesley, 2003.
- 8) N. Provos, "A Virtual Honeypot Framework," in Proc. USENIX Security Symp., San Diego, CA, USA, 2004.
- 9) H. Liao et al., "Intrusion Detection System: A Comprehensive Review," *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 16–24, 2013.
- 10) N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection," in Proc. IEEE Military Commun. Inf. Syst. Conf., Canberra, Australia, 2015, pp. 1–6.
- 11) R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion

- Detection," in Proc. IEEE Symp. Security Privacy, Oakland, CA, USA, 2010, pp. 305–316.
- 12) N. Shone et al., "A Deep Learning Approach to Network Intrusion Detection," IEEE Trans. Emerg. Topics Comput. Intell., vol. 2, no. 1, pp. 41–50, 2018.
 - 13) C. Yin et al., "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," IEEE Access, vol. 5, pp. 21954–21961, 2017.
 - 14) M. Nawrocki et al., "A Survey of Honeypot Software and Data Sources," arXiv preprint, 2016.
 - 15) S. Vyas, V. Mavroudis, and P. Burnap, "Towards the Deployment of Realistic Autonomous Cyber Network Defence: A Systematic Review," ACM Computing Surveys, vol. 58, no. 1, pp. 1–36, 2026.
 - 16) G. Sarraf and V. Pal, "Autonomous Threat Detection and Response in Cloud Security: A Comprehensive Survey of AI-Driven Strategies," arXiv preprint arXiv:2601.03303, 2026.
 - 17) M. M. Ismail et al., "Next-Generation Cybersecurity: A Deep Survey of AI and Soft Computing Techniques for Autonomous and Explainable Defense Systems," International Journal of Computers and Informatics, 2025.
 - 18) H. Alqahtani and P. Ahuja, "Secure Autonomous Cyber Defense with LLM Agents: A Systematic Review of Autonomy, Tool-Augmented Reasoning, and Governance Constraints," Computers and Electrical Engineering, vol. 135, 2026.
 - 19) Kott et al., "Autonomous Intelligent Cyber-defense Agent (AICA) Reference Architecture Release 2.0," arXiv preprint arXiv:1803.10664, 2018.
 - 20) Y. Pa et al., "IoTPOT: Analysing the Rise of IoT Compromises," in Proc. USENIX Workshop Offensive Tech