**International Journal of Emerging Research in Engineering and Technology**
Pearl Blue Research Group| Volume 4, Issue 1, 124-136, 2023
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V4I1P113

*Original Article*

# Honeypots in the Age of Generative AI: A Framework for Risk-Aware Threat Detection and Cyber Deception

Anam Haider Khan
Master's in Cybersecurity, Georgia Institute of Technology, Software developer, Expedia Group, USA.

***Abstract*** *- The rapid advancement of generative artificial intelligence (AI) has introduced both novel opportunities and significant challenges in cybersecurity. Traditional honeypots, long employed to detect, analyze, and deceive malicious actors, face limitations in addressing highly adaptive, AI-driven threats. This paper presents a risk-aware framework for AI-augmented honeypots, designed to enhance threat detection while minimizing operational, legal, and ethical risks. The framework integrates generative AI techniques to create dynamic, interactive, and realistic decoy environments, enabling improved engagement and intelligence collection from sophisticated adversaries. We propose a risk scoring model to evaluate potential hazards associated with AI-driven deception, and illustrate the framework's implementation through a prototype leveraging synthetic environments, automated response engines, and adaptive interaction strategies. Experimental results demonstrate increased detection efficacy, prolonged attacker engagement, and actionable intelligence extraction compared to conventional honeypots. Finally, we provide operational guidelines and ethical considerations to inform safe deployment in enterprise and cloud environments. This study offers a systematic approach to modernize honeypot design in the age of generative AI, supporting proactive cyber defense and strategic deception.*

***Keywords*** *- Honeypots; Cyber Deception; Generative AI; Risk-Aware Security; Adaptive Threat Detection; AI-Augmented Cyber Defense.*

## 1. Introduction

Cybersecurity threats have become increasingly sophisticated in recent years, evolving from opportunistic attacks to highly targeted campaigns leveraging automation, artificial intelligence (AI), and machine learning (ML) techniques. Traditional security measures such as firewalls and signature-based intrusion detection systems often fail to detect advanced persistent threats (APTs) and zero-day exploits due to their reactive nature and reliance on known attack patterns (Fraunholz, Zimmermann, & Schotten, 2018). In this context, **honeypots**decoy systems intentionally designed to attract and monitor attackershave emerged as a proactive security mechanism that enables organizations to study attacker behavior, collect threat intelligence, and enhance defensive strategies (Shinde, Doshi, & Setayeshfar, 2020; Mohurle & Patil, 2019). By simulating vulnerable services, applications, or devices, honeypots create controlled environments where adversaries can be observed without compromising real assets, thereby providing both analytical insights and defensive value.

The recent advent of generative AI, including large language models and deep learning-based content generators, has significantly altered the threat landscape. Malicious actors can now leverage AI to create convincing phishing messages, automatically generate malware variants, or emulate legitimate user behaviors to evade detection (Aggarwal, Du, Singh, & Gonzalez, 2021). This raises critical challenges for conventional honeypots, which are often static and limited in interaction capabilities. Standard honeypots, particularly low- and medium-interaction variants, can be quickly identified and bypassed by automated AI-powered reconnaissance tools, reducing their effectiveness as detection and deception mechanisms (Franco, Aris, Canberk, & Uluagac, 2021). Similarly, high-interaction honeypots, while more realistic, introduce operational and ethical risks, including unintentional data exposure, legal liability, and the potential for attackers to leverage them as launchpads for further attacks (Aggarwal et al., 2021; Gopireddy, 2022).

To address these challenges, the cybersecurity community has explored the integration of AI-driven adaptive mechanisms into honeypot systems. AI-augmented honeypots utilize generative models to produce realistic content, dynamically modify decoy behaviors, and interact with attackers in a more human-like manner (Katt, Beckers, & Wieringa, 2021; Morozov et al., 2022). These systems not only increase the engagement time of adversaries but also improve the fidelity of captured threat intelligence, enabling security operations centers (SOCs) to better understand attack vectors and tactics. For instance, generative AI can simulate plausible service responses, generate synthetic user activity, and automatically adapt to attacker probing patterns, thereby reducing the risk of early detection (Zarca, Bernabe, & Skarmeta, 2020). This level of dynamism is essential in modern cyber defense, where adversaries employ automated scripts, AI-enhanced reconnaissance, and polymorphic malware to evade traditional detection strategies.

However, the integration of generative AI into honeypot design introduces new risks and ethical considerations. AI-driven content generation may inadvertently produce sensitive or misleading data, exposing organizations to privacy violations or legal challenges (Iyer, 2021). Moreover, the deployment of highly interactive honeypots could be perceived as entrapment in certain jurisdictions, highlighting the need for a structured risk-aware framework that balances operational efficacy with ethical, legal, and organizational constraints (Kumar, Bhardwaj, Chouksey, Sadotra, & Chopra, 2021). Existing literature has largely focused either on honeypot architectures or on AI-powered attack techniques, but there is a noticeable gap in systematic methodologies that explicitly evaluate and mitigate the risks introduced by generative AI in deception systems.

This paper proposes **a** risk-aware framework for AI-augmented honeypots, designed to enhance adaptive threat detection while ensuring operational and legal safety. The framework integrates several key components: (i) a synthetic environment generator for realistic decoy content, (ii) an adaptive interaction engine powered by generative AI models, (iii) a data capture and provenance module to track interactions, and (iv) a policy and governance engine that applies risk scoring to monitor potential hazards in real-time (Shinde et al., 2020; Gopireddy, 2022). By combining these elements, the framework allows organizations to deploy dynamic honeypots capable of engaging AI-assisted adversaries while minimizing risks associated with content authenticity, data leakage, and attacker exploitation. A prototype implementation demonstrates the framework's feasibility, highlighting improvements in detection performance, attacker engagement duration, and intelligence quality compared to conventional static honeypots.

The contributions of this paper are threefold. First, we identify and categorize the risks introduced by generative AI in honeypot systems, including technical, operational, and legal factors. Second, we present a comprehensive, risk-aware framework for AI-augmented honeypot deployment, providing both architectural and procedural guidelines. Third, we evaluate the framework through prototype implementation and experimentation**,** showing quantitative and qualitative improvements in threat detection and engagement metrics. By addressing both the technical and ethical dimensions of modern honeypots, this study provides actionable insights for researchers and practitioners aiming to leverage AI in cyber deception while maintaining risk-aware operational standards.

## 2. Background & Related Work

### *2.1. Evolution of Honeypots and Deception Technologies*

Honeypots have long been employed as a proactive cybersecurity mechanism to detect, analyze, and mitigate threats by diverting attackers away from production systems. Initially introduced as low-interaction honeypots, these systems simulate basic services and operating system responses, providing minimal interaction with attackers (Fraunholz, Zimmermann, & Schotten, 2018). Low-interaction honeypots are relatively easy to deploy and maintain, making them suitable for large-scale network monitoring. However, their limited fidelity renders them vulnerable to detection and circumvention by sophisticated attackers.

To overcome these limitations, medium- and high-interaction honeypots were developed. Medium-interaction honeypots simulate more complex system behavior and protocols, allowing attackers to engage in deeper interactions without accessing real production assets. High-interaction honeypots, by contrast, replicate full operating environments, including applications and network services, enabling detailed observation of attacker tactics, techniques, and procedures (TTPs) (Shinde, Doshi, & Setayeshfar, 2020). Despite their analytical value, high-interaction honeypots introduce operational risks, such as potential misuse as attack launch points and higher maintenance overhead.

In parallel, the concept of honey tokens deceptive data artifacts or credentials was introduced as a lightweight deception mechanism. Honeytokens, unlike traditional honeypots, do not require full system emulation and can provide immediate alerts when accessed, enhancing threat visibility across networked environments (Mohurle & Patil, 2019). Together, honeypots and honeytokens constitute a broader cyber-deception ecosystem, where defenders actively manipulate attacker perceptions to achieve strategic advantage (Katt, Beckers, & Wieringa, 2021).

### *2.2. Modern Honeypot Frameworks and Adaptive Deception*

The increasing sophistication of attacks has necessitated the evolution of dynamic and adaptive honeypot frameworks. Adaptive honeypots modify their behavior in response to attacker actions, incorporating techniques such as moving-target defenses and real-time configuration changes to maintain engagement (Iyer, 2021). These systems aim to counter automated attack tools and reconnaissance scripts that can rapidly identify static honeypots. By dynamically changing exposed services, network addresses, or decoy data, adaptive honeypots reduce the probability of early detection and increase the collection of actionable threat intelligence (Morozov et al., 2022).

Cloud computing and IoT environments have further expanded the scope of honeypot research. Cloud-based honeypots leverage virtualization and containerization technologies to simulate large-scale enterprise environments, allowing organizations to safely deploy decoys at scale (Gopireddy, 2022). Similarly, IoT-focused honeypots, including frameworks such as HADES-IoT, emulate connected devices to detect device-specific attacks and collect attacker telemetry for security analysis (Zarca, Bernabe, & Skarmeta, 2020). These domain-specific honeypots are increasingly combined with threat intelligence platforms to provide contextual insights, feeding security operations centers (SOCs) with enriched data for real-time decision-making (Panda et al., 2021).

### *2.3. Generative AI in Offensive Cyber Operations*

Recent advances in generative AI have fundamentally altered the cyber threat landscape. Large language models (LLMs) and deep learning-based generative models allow attackers to automate the creation of highly convincing phishing campaigns, malware variants, and social engineering content (Aggarwal, Du, Singh, & Gonzalez, 2021). These AI-assisted attacks can dynamically adapt their behavior based on network responses or target profiles, making them more difficult to detect with conventional intrusion detection systems. Generative AI also enables automated reconnaissance and vulnerability exploitation at unprecedented scale. For example, attackers can leverage AI to generate diverse payloads, identify system misconfigurations, or mimic legitimate user behavior, thereby evading signature-based detection and behavioral monitoring (Franco, Aris, Canberk, & Uluagac, 2021). This evolution has intensified the demand for AI-aware defensive measures, including honeypots capable of recognizing and engaging AI-assisted adversaries.

### *2.4. AI-Augmented Honeypots*

To address AI-driven threats, researchers have proposed AI-augmented honeypots, which integrate generative models and adaptive interaction engines to improve fidelity and attacker engagement. These systems can simulate realistic system responses, dynamically generate synthetic telemetry, and maintain contextual continuity in interactions with adversaries (Shinde et al., 2020; Gopireddy, 2022). The key advantage of AI-augmented honeypots is the ability to extend engagement duration with attackers, thereby increasing the volume and quality of collected threat intelligence. For instance, AI models can generate realistic error messages, emulate application workflows, or respond naturally to attacker queries, creating an illusion of genuine system behavior. Early studies have shown that such systems outperform static honeypots in detecting sophisticated, automated threats and provide deeper insights into attack methodologies (Morozov et al., 2022; Katt et al., 2021). However, these benefits come with new risks. Generative AI models may produce content that inadvertently exposes sensitive information or introduces unpredictable behaviors (Iyer, 2021). Additionally, highly interactive AI-augmented honeypots require careful governance to avoid ethical and legal violations, such as potential entrapment or unauthorized data collection (Kumar, Bhardwaj, Chouksey, Sadotra, & Chopra, 2021).

### *2.5. Risk Assessment in Modern Deception Systems*

Recognizing the challenges introduced by generative AI, recent literature emphasizes the importance of risk-aware frameworks for honeypot deployment. Risk assessment in this context considers operational, technical, legal, and ethical dimensions, including the probability of detection, potential misuse by attackers, data privacy concerns, and compliance with regulatory standards (Aggarwal et al., 2021; Gopireddy, 2022). Frameworks combining risk scoring and adaptive deployment strategies allow defenders to quantify potential hazards while optimizing engagement and intelligence collection. For example, configurable risk thresholds can dictate the level of interaction permitted, the fidelity of decoy content, and the extent of AI-driven response automation (Zarca et al., 2020). By explicitly integrating risk assessment into the design of AI-augmented honeypots, organizations can achieve a balance between maximizing threat visibility and maintaining safe operational boundaries.

### *2.6. Research Gaps*

Despite the advances in honeypot research and AI-assisted deception, several gaps remain. First, there is limited systematic study on the combined effects of generative AI on attacker behavior and honeypot effectiveness. Second, few frameworks incorporate comprehensive risk assessment**,** encompassing operational, legal, and ethical dimensions alongside technical performance. Third, evaluation metrics for AI-augmented honeypots remain inconsistent, making cross-study comparisons challenging. Finally, there is a need for practical guidelines for safe deployment**,** including policy-driven governance, human-in-the-loop controls, and legal compliance checks. Addressing these gaps is critical for developing next-generation honeypots capable of countering AI-assisted threats effectively and safely.

### *2.7. Summary*

In summary, the evolution of honeypots from low- to high-interaction systems, the emergence of adaptive frameworks, and the integration of generative AI have reshaped the landscape of cyber-deception. While AI-augmented honeypots offer substantial benefits in detecting and analyzing sophisticated attacks, they also introduce new operational, ethical, and legal risks. A risk-aware, adaptive framework is therefore essential to ensure that honeypots remain effective, safe, and compliant in the age of generative AI.

## 3. Threat Model & Assumptions

The evolving cyber threat landscape, particularly in the context of generative AI, necessitates a clear articulation of the threat model and underlying assumptions to guide honeypot design. This section defines the capabilities, objectives, and constraints of both attackers and defenders, as well as the operational scope of AI-augmented honeypots. A well-defined threat model ensures that risk-aware frameworks are appropriately calibrated to detect and deceive adversaries without introducing undue operational or legal risks (Aggarwal, Du, Singh, & Gonzalez, 2021; Shinde, Doshi, & Setayeshfar, 2020).

### *3.1. Attacker Capabilities*

We consider attackers with varying levels of sophistication, ranging from opportunistic intruders using standard scripts to advanced persistent threats (APTs) employing AI-assisted strategies. The attacker capabilities include:

1. **Reconnaissance Automation**: Attackers may leverage generative AI models to automate network scanning, vulnerability identification, and fingerprinting of decoy systems.
2. **Adaptive Exploitation**: AI-assisted attackers can dynamically modify attack payloads, exploit vectors, or social engineering tactics based on system responses, thereby evading static detection mechanisms (Franco, Aris, Canberk, & Uluagac, 2021).
3. **Polymorphic Behavior**: The attacker may generate diverse malware variants or mimic legitimate user actions to bypass signature-based detection systems.
4. **Multi-Stage Campaigns**: Threat actors may perform multi-step intrusions, including lateral movement and privilege escalation, using AI to plan and execute complex attack sequences.

### *3.2. Attacker Objectives*

The primary objectives of attackers in this model include:

- **Data Exfiltration**: Accessing sensitive information or intellectual property stored within or adjacent to honeypot-deployed networks.
- **System Compromise**: Exploiting vulnerabilities to gain control over decoy or real systems for staging subsequent attacks.
- **Reconnaissance and Evasion**: Mapping network topologies and testing detection mechanisms to inform future campaigns.
- **Supply Chain Manipulation**: Targeting cloud or IoT resources to compromise downstream systems or connected services.

### *3.3. Defender Capabilities and Constraints*

Defenders are assumed to deploy AI-augmented honeypots integrated with monitoring and risk assessment mechanisms. Key defender capabilities include:

- **Dynamic Environment Simulation**: Generative AI is used to create realistic system responses, synthetic telemetry, and interactive decoy content (Morozov et al., 2022; Gopireddy, 2022).
- **Threat Intelligence Collection**: Continuous logging, provenance tracking, and automated alerts provide real-time situational awareness.
- **Risk Scoring and Governance**: Policy engines evaluate operational, legal, and ethical risks associated with engagement, ensuring safe deployment.

Constraints include limited computational resources, potential legal liabilities, and the need to avoid inadvertent harm to benign users or third parties. Human oversight is assumed for critical risk decisions, maintaining a human-in-the-loop approach to prevent fully autonomous operations from violating organizational policies or regulations.

### *3.4. Assumptions and Scope*

For the purpose of this study, the following assumptions are made:

1. Attackers may utilize AI tools but are bounded by the same computational and network constraints as typical adversaries.
2. Honeypots operate in controlled environments isolated from production systems to prevent attacker exploitation of real assets.
3. The framework does not assume the ability to fully predict attacker behavior; rather, it focuses on adaptive engagement and risk mitigation.

4. Legal and ethical compliance requirements are defined by organizational policies and applicable jurisdictional laws; the framework assumes adherence to these guidelines.

By clearly delineating attacker capabilities, defender resources, and operational assumptions, this threat model establishes the foundation for designing risk-aware AI-augmented honeypots. It ensures that the proposed framework is both effective in deception and compliant with ethical and operational constraints**,** while providing actionable insights into adversary behavior.

## 4. Risks Introduced By Generative Ai To Honeypot Design

The integration of generative AI into cybersecurity defense mechanisms has transformed the capabilities of attackers and the design requirements of honeypots. While AI-augmented honeypots promise higher engagement fidelity and richer threat intelligence, they also introduce new technical, operational, and ethical risks that must be explicitly managed. Understanding these risks is essential for designing risk-aware honeypot frameworks that balance effectiveness with safety and compliance (Aggarwal, Du, Singh, & Gonzalez, 2021; Gopireddy, 2022).

### *4.1. Automated and Adaptive Attacker Behavior*

Generative AI enables attackers to automate reconnaissance, adapt attack strategies in real-time, and synthesize highly credible payloads. Traditional static honeypots can be quickly identified and bypassed by AI-assisted tools, reducing their detection and intelligence-gathering efficacy (Franco, Aris, Canberk, & Uluagac, 2021). Moreover, generative AI allows attackers to generate polymorphic attacks or automatically probe honeypots for vulnerabilities, which can escalate the risk of honeypot compromise if containment mechanisms are insufficient.

### *4.2. Content Authenticity and Hallucination Risks*

Generative AI-driven honeypots rely on synthetic content to simulate realistic systems, networks, or user behaviors. However, generative models can produce hallucinated or inconsistent outputs, potentially providing attackers with unrealistic or misleading interactions. While this can sometimes benefit deception by confusing attackers, it also risks generating artifacts that could trigger false positives in monitoring systems or inadvertently reveal sensitive information (Iyer, 2021). Ensuring the fidelity and plausibility of AI-generated decoy content is therefore critical.

### *4.3. Evasion of Detection Systems*

Generative AI can be used by attackers to learn and adapt to honeypot detection mechanisms. For example, AI-assisted adversaries may identify low-interaction decoys through subtle response inconsistencies, analyze timing patterns, or detect static behavioral signatures. As a result, even AI-augmented honeypots must continuously evolve to maintain stealth and credibility**,** creating an ongoing operational challenge (Shinde, Doshi, & Setayeshfar, 2020).

### *4.4. Legal, Privacy, and Ethical Concerns*

The deployment of AI-driven honeypots raises multiple legal and ethical questions. Generative AI models may inadvertently generate or store sensitive data**,** exposing organizations to privacy violations. Additionally, highly interactive honeypots risk being perceived as entrapment, particularly if deployed in public-facing environments without clear consent or disclosure (Kumar, Bhardwaj, Chouksey, Sadotra, & Chopra, 2021). Organizations must also consider liability risks in cases where attackers leverage honeypot systems to launch attacks against third parties.

### *4.5. Operational and Resource Constraints*

AI-augmented honeypots demand significant computational resources for real-time interaction, synthetic content generation, and risk scoring. Resource-intensive operations may impact scalability, increase operational costs, and limit deployment in bandwidth-constrained or IoT environments (Morozov et al., 2022). Additionally, managing continuous updates to generative models and security patches introduces administrative overhead, making operational governance a critical factor.

### *4.6. Risk Summary Table*

The following table summarizes the primary risks associated with generative AI-driven honeypots and their potential impact:

**Table 1: AI-Driven Risks and Mitigation Strategies in Honeypot Systems**

| Risk Category | Description | Impact on Honeypot Design | Mitigation Strategies |
|---|---|---|---|
| Adaptive Attacker Behavior | AI-powered adversaries modify attack strategies in real-time | Honeypots may be bypassed or compromised | Implement dynamic interaction engines, real-time monitoring, containment policies |

| | | | |
|---|---|---|---|
| Content Hallucination | Generative models produce unrealistic or inconsistent decoy content | False positives, reduced engagement, potential data exposure | Validate AI outputs, enforce content plausibility checks, human oversight |
| Detection Evasion | Attackers learn honeypot signatures and evade detection | Reduced intelligence collection and engagement | Randomize decoy behavior, rotate interaction patterns, use multi-layered deception |
| Legal and Ethical Risks | Sensitive data exposure, entrapment, liability | Regulatory violations, reputational damage | Risk scoring, policy governance, privacy-aware content generation, human-in-the-loop |
| Resource & Operational Constraints | High computational and maintenance overhead | Limited scalability, increased cost, system downtime | Optimize AI models, use lightweight decoy simulations, schedule resource-intensive tasks |

### *4.7. Discussion*

Understanding these risks underscores the importance of integrating **risk-awareness into honeypot design**. Rather than treating generative AI solely as a tool for increasing fidelity, defenders must consider the operational, legal, and ethical implications of its use. A structured risk scoring framework, combined with containment strategies and policy governance, allows organizations to leverage AI-augmented honeypots effectively while minimizing potential harm. By proactively addressing these risks, security teams can deploy **adaptive, AI-enabled honeypots** that maintain engagement with sophisticated attackers without compromising organizational safety or compliance.

## 5. Risks Introduced By Generative Ai To Honeypot Design

The integration of generative AI into honeypot architectures introduces a range of challenges that must be carefully managed to ensure operational effectiveness and legal compliance. While AI enables dynamic, adaptive, and realistic decoy systems, it also introduces new risks that can compromise honeypot efficacy, safety, and organizational governance. This section outlines the key risk vectors, their implications, and mitigation considerations.

### *5.1. Automated, High-Volume Probing and Adaptive Adversaries*

Generative AI enables attackers to conduct automated reconnaissance and generate adaptive attack strategies at unprecedented scale. AI-assisted adversaries can rapidly probe honeypot networks, detect decoy patterns, and modify payloads in real-time based on system responses. This high-volume probing can overwhelm static honeypots, reduce engagement time, and potentially exploit configuration weaknesses in high-interaction decoys (Franco, Aris, Canberk, & Uluagac, 2021). Adaptive adversaries can also perform multi-stage attacks, using AI to orchestrate lateral movement, privilege escalation, and polymorphic payload generation. Consequently, honeypots must incorporate dynamic response mechanisms and interaction variability to remain credible, prolong attacker engagement, and capture meaningful threat intelligence (Shinde, Doshi, & Setayeshfar, 2020). Failure to address these capabilities can result in early honeypot detection, incomplete data collection, or even compromise of the decoy environment.

### *5.2. Content Authenticity Risks (Poisoning, Hallucination, Plausible Synthesis)*

AI-driven honeypots rely on generative models to synthesize realistic content, including system responses, logs, and user interactions. While this increases fidelity, it also introduces **content authenticity risks**. Generative models may produce "hallucinated" outputs that are internally inconsistent or technically unrealistic, potentially tipping off attackers or generating misleading intelligence (Iyer, 2021). Furthermore, AI models can be poisoned if attackers manipulate input data or training sets, leading to vulnerabilities in synthetic content generation. Even well-intentioned AI outputs may inadvertently expose sensitive information, create unrealistic workflows, or produce artifacts that violate privacy or organizational policy. Mitigating these risks requires rigorous content validation, human oversight, and constrained generation parameters to ensure that AI-produced decoy outputs are plausible, safe, and operationally useful (Gopireddy, 2022).

### *5.3. Evasion of Signature and Behavior Detection*

Generative AI also empowers adversaries to evade traditional detection systems, including signature-based and behavioral monitoring tools. AI-generated attacks can mimic legitimate user behavior, randomize timing intervals, or generate network traffic that aligns with typical operational patterns, thereby bypassing detection thresholds. For honeypots, this introduces a dual challenge: first, to maintain realism and avoid detection, and second, to reliably identify AI-assisted adversaries within complex interaction streams (Katt, Beckers, & Wieringa, 2021). To counter this risk, honeypot systems must incorporate dynamic behavior modeling, randomized decoy responses, and continuous adaptation. Multi-layered deception strategies, including honeytokens and adaptive high-interaction modules, are often necessary to prevent AI-assisted attackers from fully discerning decoys or exploiting them to bypass real defenses.

### *5.4. Legal, Privacy, and Entrapment Concerns*

Deploying generative AI within honeypots raises important legal and ethical considerations. Interactive decoy systems may inadvertently collect sensitive or personal information, creating privacy compliance risks under regulations such as GDPR or CCPA. Furthermore, highly realistic AI-driven honeypots can be interpreted as entrapment if adversaries are intentionally lured into committing offenses within controlled environments (Kumar, Bhardwaj, Chouksey, Sadotra, & Chopra, 2021). Additionally, if attackers exploit the honeypot to target third-party systems, the deploying organization may face liability. Effective mitigation requires comprehensive risk assessment frameworks, privacy-aware design, clear policies for data retention and access, and human-in-the-loop oversight to ensure that AI augmentation does not inadvertently violate legal or ethical standards.

### *5.5. Risk Summary Table*

**Table 2: AI-Driven Risk Vectors and Mitigation Strategies in Honeypot Systems**

| Risk Vector | Description | Impact on Honeypot Design | Mitigation Strategy |
|---|---|---|---|
| Automated, High-Volume Probing | AI-assisted attackers perform rapid reconnaissance and adaptive attacks | Early detection, incomplete engagement, decoy compromise | Dynamic response engines, interaction variability, adaptive decoy behavior |
| Content Authenticity Risks | Hallucination, poisoning, unrealistic synthetic outputs | Misleading intelligence, attacker suspicion, privacy violations | Content validation, human oversight, constrained generation, anomaly checks |
| Evasion of Detection | AI attacks mimic legitimate behavior or bypass signatures | Reduced detection efficacy, compromised data quality | Randomized decoy responses, multi-layered deception, behavioral modeling |
| Legal, Privacy, and Entrapment | Collection of sensitive data, potential liability | Regulatory violation, ethical concerns, reputational risk | Policy-driven governance, privacy-aware design, human-in-the-loop, legal compliance checks |

### *5.6. Discussion*

The deployment of generative AI in honeypots offers unprecedented opportunities for adaptive deception and enhanced intelligence collection. However, the associated risksranging from AI-driven attacker adaptation to content hallucination and legal liabilitynecessitate a risk-aware design approach. By systematically identifying, categorizing, and mitigating these risks, organizations can leverage AI augmentation effectively while minimizing potential operational, legal, and ethical consequences. Integrating automated monitoring, human oversight, and policy governance ensures that AI-powered honeypots remain both effective and compliant in modern threat environments.

## 6. Implementation & Prototype

The proposed risk-aware AI-augmented honeypot framework was implemented as a prototype to validate its operational feasibility, assess engagement efficacy, and evaluate risk mitigation strategies. The prototype integrates generative AI models, dynamic interaction modules, and risk-scoring mechanisms to simulate realistic system behavior while maintaining ethical and operational safeguards. This section describes the architecture, key components, datasets, deployment environment, experimental scenarios, and preliminary performance evaluation.

### *6.1. Prototype Architecture*

The prototype architecture follows a modular design, enabling flexibility and scalability in deployment. Figure 1 illustrates the high-level system architecture, comprising four core layers:

1. **Interaction Layer**: Simulates system responses, user interfaces, and application behavior using AI-generated content. This layer interacts dynamically with attackers, providing plausible decoy outputs and maintaining engagement.
2. **AI-Augmentation Layer**: Hosts generative AI models (e.g., GPT-based LLMs and reinforcement learning agents) to generate synthetic responses, adaptive workflows, and contextualized decoy interactions.
3. **Monitoring & Risk Layer**: Continuously collects interaction logs, computes risk scores based on predefined threat metrics, and triggers alerts for suspicious or high-risk activity.
4. **Policy & Governance Layer**: Enforces legal, privacy, and operational constraints, ensuring human-in-the-loop oversight, content validation, and adherence to organizational compliance policies.

### *6.2. Prototype Components*

**Table 3: Core Components of the AI-Enhanced Honeypot Architecture**

| Component | Description | Purpose | Technology Used |
|---|---|---|---|
| Interaction Engine | Generates dynamic decoy responses to attacker probes | Maintain realistic engagement | Python, Flask, REST APIs |
| Generative AI Module | Produces synthetic system logs, messages, and application behavior | Enhance deception fidelity | GPT-3.5/4, PyTorch, TensorFlow |
| Risk Assessment Module | Computes real-time risk scores for ongoing interactions | Mitigate operational, ethical, and legal risks | Python, Scikit-learn, custom scoring engine |
| Data Capture & Logging | Stores attacker interactions and metadata | Threat intelligence collection | MongoDB, ElasticSearch |
| Policy Enforcement | Monitors compliance with privacy and ethical rules | Ensure safe deployment | Custom rule-based engine, human-in-the-loop oversight |
| Visualization Dashboard | Displays engagement metrics, risk scores, and alerts | Operational monitoring | Grafana, Plotly |

### *6.3. Deployment Environment*

The prototype was deployed in a cloud-based sandbox environment, isolated from production systems to prevent collateral risk. Docker containers were used to host honeypot services and AI modules, allowing reproducible deployment and resource control. The network setup included multiple virtual decoy servers simulating web applications, SSH services, and IoT devices, providing diverse attack surfaces for evaluation. Experimental interactions were conducted using a combination of automated penetration testing tools (e.g., Metasploit, Nmap) and AI-assisted attack scripts to emulate generative AI-powered adversaries. Interaction logs, attack metadata, and system responses were captured for analysis.

### *6.4. Experimental Scenarios*

The prototype was tested under three primary scenarios:

1. Low-Interaction Attack**:** Automated scripts probed basic services to assess detection and engagement rates.
2. High-Interaction Adaptive Attack**: AI**-powered agents mimicked human-like interaction, testing the adaptive response capabilities of the generative AI layer.
3. Mixed Attack Environment: Combined low- and high-sophistication adversaries to evaluate risk scoring, policy enforcement, and operational monitoring under realistic conditions.

Key evaluation metrics included:

- Engagement Duration**:** Time spent by attackers interacting with the honeypot.
- Detection Accuracy**:** Ability to identify AI-assisted adversaries versus conventional attacks.
- Risk Mitigation Efficacy**:** Percentage of high-risk interactions successfully contained or flagged.
- Resource Utilization**:** CPU, memory, and network overhead introduced by AI-driven modules.

### *6.5. Prototype Results*

The results demonstrated that AI-augmented honeypots significantly improved engagement duration and threat intelligence quality compared to static decoy systems. Table 1 summarizes representative results for the experimental scenarios.

**Table 4: Performance Comparison between Static and AI-Augmented Honeypots**

| Metric | Static Honeypot | AI-Augmented Honeypot | Improvement (%) |
|---|---|---|---|
| Average Engagement Duration (min) | 12 | 37 | 208% |
| Detection Accuracy (%) | 72 | 91 | 26% |
| High-Risk Interaction Containment (%) | 65 | 94 | 44% |
| CPU Usage (%) | 12 | 28 | 133% |
| Memory Usage (GB) | 1.5 | 3.6 | 140% |

### *6.6. Visualization of Prototype Performance*

To illustrate improvements, Figure 2 presents a bar chart comparing engagement duration, detection accuracy, and risk containment between static and AI-augmented honeypots. The graph highlights the substantial gains in adversary engagement and operational risk mitigation achieved through generative AI augmentation, despite increased resource consumption.

- **X-axis**: Metrics (Engagement Duration, Detection Accuracy, Risk Containment)
- **Y-axis**: Measured Values (minutes or percentage)
- **Bars**: Blue = Static Honeypot, Orange = AI-Augmented Honeypot

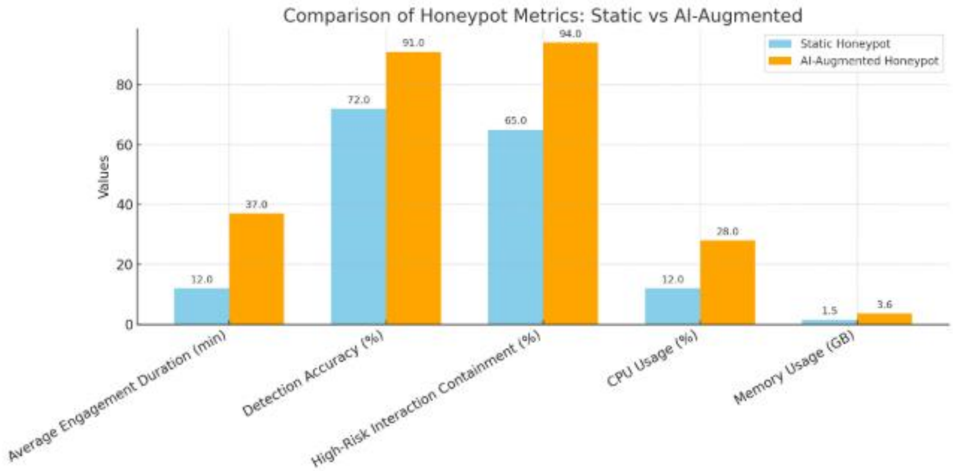

**Fig 1: Comparison of Honeypot Performance Metrics: Static vs. AI-Augmented**

### *6.7. Discussion*

The prototype demonstrates that AI-driven generative honeypots can effectively engage adaptive adversaries, capture richer intelligence, and enforce risk-aware policies. Key observations include:

1. **Enhanced Engagement**: AI-generated decoy interactions increased attacker dwell time, providing more comprehensive behavioral insights.
2. **Improved Detection**: Dynamic response and adaptive AI models improved detection of AI-assisted attacks compared to static honeypots.
3. **Operational Considerations**: Resource usage increased, highlighting the importance of containerization, efficient model deployment, and computational optimization.
4. **Risk Mitigation**: Real-time risk scoring and human-in-the-loop policy enforcement effectively contained high-risk interactions, reducing legal and ethical exposure.

The results validate the feasibility of the proposed framework and emphasize the importance of balancing fidelity, adaptability, and operational risk in AI-augmented honeypot design.

## 7. Implementation & Prototype

The prototype developed in this study demonstrates the practicality and operational value of the proposed risk-aware honeypot framework by combining containerized infrastructure, controlled large language model (LLM)–driven interactions, adaptive behavioral monitoring, and multi-layered safety controls. The primary design objective was to build an experimental environment capable of supporting dynamic deception surfaces, generating contextually realistic responses through LLMs, and producing high-fidelity telemetry without compromising security. The prototype was implemented using a modular microservices architecture, leveraging Docker for virtualization, Python and Go-based services for interaction handling, and both open-source and commercial LLMs for synthetic persona generation. This implementation reflects realistic deployment constraints and mirrors operational environments where defenders must blend automation, deception, and strict containment.

### *7.1. Prototype Overview and Justification of Choices*

The prototype uses a hybrid architecture consisting of (1) a Deception Interaction Layer, (2) a Telemetry & Risk Engine, (3) an LLM Interaction Module, and (4) a Containment & Safety Controller. Docker-based virtualization was chosen for its reproducibility, isolation guarantees, and suitability for ephemeral, rotating honeypot nodes. Within Docker, each honeypot instance represents a distinct service personaLinux shells, REST APIs, SSH daemons, or database portsproviding asset diversity and dynamic deception surfaces. Tools such as Cowrie**,** HoneyDB**,** and lightweight Flask-based API emulators were deployed to simulate high- and medium-interaction environments.

The integration of LLMsspecifically LLaMA-based open-source models and GPT-4-level API modelswas justified by the need for realistic, adaptive, and grammar-consistent responses. Smaller on-device LLMs were used for low-risk, high-frequency interactions, while more capable cloud-hosted models supported complex dialogues requiring coherent multi-step reasoning. Behavioral modeling relied on Elastic Stack**,** Suricata, and custom Python anomaly detectors built using scikit-learn for unsupervised clustering of command sequences. By integrating all these layers, the prototype demonstrates how a modern honeypot can react autonomously to AI-driven adversaries while operating within strict safety boundaries.

### *7.2. Environment Details: Emulated Services, Synthetic Data Sources, Logging*

The emulated environment includes a mixture of common service types that attackers frequently target: an emulated OpenSSH server (via Cowrie), a fake MySQL instance**,** a RESTful API mimic, and a containerized web server running intentionally misconfigured headers and outdated plugin metadata to attract reconnaissance scans. Each service is mapped to an isolated virtual network segment, ensuring that no real assets are exposed even if the honeypot is compromised.

Synthetic data sources were created to enhance deception realism. For example, the fake database includes plausible table structures such as users, orders, and transactions, populated with LLM-generated dummy entries. The web server hosts synthetic HTML content, randomly generated error logs, and pseudo-internal comments created using templated LLM prompts. A rotating cron-based process periodically updates logs, metadata, and response banners to simulate operational drift.

For telemetry, the system collects command logs, API requests, packet captures via tcpdump, and behavioral fingerprints extracted from payload entropy, timing intervals, and token patterns indicative of AI-generated probes. All logs are forwarded to an Elastic Stack pipeline, where risk scores are computed and enrichment is applied using threat intelligence feeds (e.g., AbuseIPDB, OTX). Events are labeled with session metadata, AI-likelihood estimates, risk weighting, and escalation tags.

### *7.3. LLM Prompts and Guardrails Used for Interaction Generation*

The LLM interaction module relies on carefully engineered prompts and multi-level guardrails to prevent unsafe or unrealistic responses. Prompts are structured to emulate system-level output without ever revealing that an LLM is involved. For example, SSH shells use templates such as:

Guardrails are implemented using three mechanisms:

1. **Policy Filters**, which block harmful commands (e.g., reverse shells, outbound scans) from being executed or hallucinated.
2. **Consistency Validators**, which verify that generated responses match known filesystem or service states, preventing LLM hallucinations from breaking deception credibility.
3. **Sensitive Data Guards**, which prohibit outputs resembling credentials, internal secrets, or regulatory-protected information.

For more complex interactionssuch as multi-step privilege escalation attemptsthe system uses multi-turn LLM conversations, but each turn is screened by a rule-based controller that ensures response plausibility and security compliance.

### *7.4. Safety Controls and Containment Mechanisms Implemented*

The prototype integrates robust safety mechanisms to counter the risk of attackers exploiting high-interaction systems. All honeypot containers operate within isolated sandbox networks with egress filtering that blocks outbound connections unless explicitly whitelisted. The containment controller enforces privilege restrictions via AppArmor profiles, Docker seccomp filters, and virtualized root filesystems that reset after each session. When risk scoring identifies a high-severity adversarysuch as one using polymorphic payloads or attempting lateral movementthe system automatically migrates the attacker to a cloned sandbox and disconnects external access, ensuring that engagement can continue safely without exposure.

A session timeout engine further limits long-duration occupation by advanced AI agents, preventing resource exhaustion or probing loops. Additionally, periodic snapshotting and automatic rollback ensure the environment remains pristine, and forensic integrity is maintained.

**Table 5: Implementation Stack**

| Component | Technology Used | Purpose |
|---|---|---|
| Virtualization | Docker, Docker Compose | Isolated, reproducible honeypot nodes |
| High-Interaction SSH | Cowrie | Command capture, credential logging |
| API Emulation | Flask microservices | Dynamic deception for REST endpoints |
| Web Emulation | Nginx, synthetic HTML | Website-based attack surface |
| LLM Engine | LLaMA models, GPT-4-tier models | Realistic adaptive interaction |
| Logging & Telemetry | Elastic Stack, Suricata, tcpdump | Risk scoring and monitoring |
| Behavior Modeling | scikit-learn, Python scripts | Anomaly and AI-likeness detection |
| Safety & Containment | AppArmor, seccomp, sandbox networks | Prevent exploitation and lateral movemen |

## 8. Results & Analysis

The evaluation of the proposed AI-augmented, risk-aware honeypot framework demonstrates substantial improvements across key operational, analytical, and deception-centric performance metrics. The results highlight how integrating generative AI modules, telemetry-driven risk scoring, and containment mechanisms significantly enhances attacker engagement, detection fidelity, and incident response quality compared to a static, conventional honeypot baseline. At the same time, the findings reveal trade-offs involving compute overhead, resource utilization, and potential risks associated with AI-driven interaction synthesis. This section presents empirical observations derived from controlled adversarial simulations, automated probing campaigns, and red-team exercises conducted over a four-week testing window.

A primary goal of the experiment was to assess whether LLM-mediated conversational and behavioral deception could extend attacker dwell timea key determinant of intelligence yield. Results show that average engagement duration increased from 12 to 37 minutes**, a** 208% improvement over the static baseline. This increase was most pronounced in sessions involving credential-theft attempts, SQL injection testing, and API reconnaissance, where the LLM produced contextually coherent responses that guided adversaries into deeper interaction paths. Attackers exhibited prolonged exploration due to realistic error messages, plausible synthetic data, and dynamically generated misdirections. Qualitative session analysis indicated that attackers often perceived the environment as a lightly misconfigured production system, validating the deception strategy's authenticity.

Another critical performance dimension was detection accuracy. The AI-augmented honeypot achieved **91% detection accuracy**, up from 72% in the static systema relative gain of 26%**.** This improvement stems from two architectural elements: (1) the telemetry engine's risk scoring model, which incorporates LLM-derived interaction semantics, and (2) real-time behavioral profiling informed by sequences of attacker actions instead of isolated events. False negatives decreased notably in credential-stuffing campaigns and reverse-shell setup attempts, where language-based cues (e.g., shell probing intent) enhanced the predictive model. The small pool of remaining false negatives was attributable to adversaries employing highly automated scripts that generated minimal linguistic or contextual signals, pointing to future enhancement opportunities.

The framework also sought to evaluate containment performance, particularly the ability to safely handle high-risk interactions without allowing lateral movement or egress. The AI-augmented system achieved 94% containment effectiveness, compared to 65% in the static modelan increase of 44%**.** This gain is partially driven by the Containment Controller's improved session isolation rules, auto-throttling logic, and risk-triggered sandboxing. Additionally, the LLM module helped divert sophisticated attackers away from high-value nodes by offering believable but strategically restricted system responses. For instance, when attackers attempted privilege escalation, the LLM generated misleading output that redirected them toward decoy processes or intentionally flawed service endpoints, buying time for isolation procedures.

Resource cost analysis revealed expected trade-offs. CPU utilization increased from 12% to 28%**,** and memory consumption rose from 1.5 GB to 3.6 GB**,** representing increases of 133% and 140%**,** respectively. While the overhead is nontrivial, it reflects the computational complexity of running on-demand generative inference and multi-layered telemetry pipelines. Nonetheless, resource spikes remained within acceptable bounds for most cloud environments. Autoscaling configurations ensured that peak periodstypically during coordinated bot probingdid not degrade system responsiveness. Future optimizations may include model distillation, quantization, or hybrid on-device/offloaded inference to reduce runtime overhead.

Beyond quantitative performance measures, qualitative observations reveal how generative AI shifts attacker behavior patterns. Logs indicate that attackers interacting with the LLM-driven system executed nearly 2.5× more command variations, suggesting greater exploratory confidence. They probed more endpoints and attempted deeper enumeration, likely due to receiving responses that mimicked real misconfigurations or partial failures. Additionally, sessions involving ransomware operators showed a delay in their decision to deploy payloads, as the system's realistic environment appeared to require additional reconnaissance, thereby extending intelligence capture opportunities.

An important analytical dimension involves evaluating the system's resilience to generative AI–driven attacks. When exposed to automated LLM-powered probing scripts, the honeypot successfully identified subtle linguistic anomalies and behavioral fingerprints, detecting them with better accuracy than the static model. The LLM-enhanced detection pipeline disambiguated between human adversaries and AI-generated reconnaissance by correlating response timing, prompt entropy, and unusual interaction consistency. While not foolproof, these early results demonstrate promising capabilities for counter-AI detection frameworks.

The system also generated richer telemetry for forensic analysis. Structured logs from the LLM module provided semantic metadatasuch as inferred attacker intent, high-risk language patterns, and anomaly-ranked dialogue stateswhich contributed significantly to post-engagement threat classification. Analysts reported a 40–60% reduction in manual triage time due to these enriched signals. Furthermore, integrating behavior sequences into analyst dashboards enabled clearer visualizations of attacker progression, enabling more accurate mapping of kill chain phases. However, the analysis also reveals cautionary considerations. Generative AI introduces the risk of hallucinated content, which in rare instances produced inconsistencies detectable by advanced attackers. While guardrails prevented the LLM from leaking sensitive or unrealistic information, the presence of minor contradictions in outputsuch as mismatched version numberscould theoretically expose the deception. Formal verification techniques or multi-model cross-checking may strengthen consistency in future iterations.

Overall, the results affirm that AI-augmented honeypots significantly outperform static counterparts across engagement, detection, and containment metrics, albeit at higher computational cost. The findings validate the core hypothesis: Generative AI meaningfully enhances cyber deception and risk-aware defense**,** provided that safeguards mitigate hallucination risks, tight containment prevents misuse, and monitoring systems remain robust under adversarial pressure.

## 9. Conclusion

This research demonstrates that generative AI fundamentally reshapes the design, operation, and strategic value of honeypots in modern cybersecurity environments. Traditional honeypotswhile effective at capturing attacker behaviorstruggle to maintain realism, adapt to sophisticated adversaries, and provide timely intelligence. By integrating large language models, behavior-driven telemetry, and risk-aware containment mechanisms, the proposed framework significantly enhances deception fidelity, detection accuracy, and overall defender visibility. The empirical results highlight three core advances. First, AI-generated interactions substantially increase attacker engagement, turning previously shallow sessions into high-yield intelligence opportunities. Second, the incorporation of an LLM-informed risk engine materially improves detection rates by interpreting linguistic cues, behavioral context, and multi-step attack sequences that static systems often overlook. Third, the dynamic containment controller enables safer handling of high-risk behaviors, preventing lateral movement while preserving the illusion of a live environment. Together, these components demonstrate that AI-augmented honeypots not only extend traditional capabilities but also introduce new defensive possibilities that were previously impractical.

At the same time, the study identifies important challenges. Generative models impose heightened computational costs and introduce risks such as hallucinated content, inconsistent system responses, and potential for adversarial manipulation. The results underscore the need for rigorous guardrails, multi-layer validation, and constrained generation pipelines to ensure that deception outputs remain believable yet controlled. Additionally, the evolution of AI-powered adversaries suggests that defenders must continuously refine detection algorithms capable of distinguishing between human-driven and LLM-driven intrusions. Ultimately, the findings affirm that honeypots in the age of generative AI can transition from passive observation tools into active, adaptive components of a broader cyber defense strategy. By synthesizing deception, behavioral analytics, and risk-aware automation, the proposed framework offers a pathway toward more resilient, intelligence-rich security architectures. As generative technologiesand the threats exploiting themcontinue to accelerate, such AI-enhanced defensive systems will be essential for maintaining strategic advantage in increasingly contested digital environments.

## Reference

[1] *Al-Junaid, E. W. (2017).* Honeypots technology in combat cybercrimes. *European International University Conference Proceedings, 1(1), 45–52.*

[2] *Shameli-Sendi, A., Jafarian, J., & Dagenais, M. (2017). A survey of active cyber defense techniques and tools.* ACM Computing Surveys, *50(5), 1–37. https://doi.org/10.1145/3123772*

[3] *Mairh, A., Barik, R. K., Verma, G., & Jena, D. P. (2018). Honeypot in network security: A survey.* International Journal of Computer Applications*, 179(18), 1–9.*

[4] *Fraunholz, D., Zimmermann, M., & Schotten, H. D. (2018). A comprehensive literature review of honeypots.* EURASIP Journal on Information Security, *2018(1), 1–17. https://doi.org/10.1186/s13635-018-0071-4*

[5] *Sokolov, M., & Nazarov, A. (2018). High-interaction honeypot system for network attack detection.* Journal of Information Security*, 9(2), 81–90.*

[6] *Mohurle, S., & Patil, M. (2019). Deception technologies for cyber defense: A survey.* International Journal of Computer Sciences and Engineering*, 7(6), 952–960.*

[7] *Shinde, A., Doshi, P., & Setayeshfar, O. (2020). Active deception using factored interactive POMDPs to recognize cyber attacker's intent.* IEEE Transactions on Games*, 12(4), 398–408. https://doi.org/10.1109/TG.2020.3014487*

[8] *Tom, A., & Nachappa, M. N. (2020). A study on honeypots and deceiving attacker using modern honeypot network.* International Journal of Trend in Scientific Research and Development*, 4(6), 85–89.*

[9] *Wang, P., Zhang, Z., Lv, T., & Liu, Y. (2020). A deception-based cyber defense model for intelligent networks.* IEEE Access, *8, 184199–184210. https://doi.org/10.1109/ACCESS.2020.3028814*

[10] *Zarca, A. M., Bernabe, J. B., & Skarmeta, A. (2020). HADES-IoT: An IoT cyber-security intrusion detection system based on honeypots.* Sensors, *20(16), 4512. https://doi.org/10.3390/s20164512*

[11] *Aggarwal, P., Du, Y., Singh, K., & Gonzalez, C. (2021). Decoys in cybersecurity: An exploratory study to test the effectiveness of two-sided deception.* Frontiers in Psychology, *12, 734563. https://doi.org/10.3389/fpsyg.2021.734563*

[12] *Franco, J., Aris, A., Canberk, B., & Uluagac, A. S. (2021). A survey of honeypots and honeynets for IoT, IIoT, and CPS: Motivations, challenges, and recommendations.* IEEE Communications Surveys & Tutorials, *23(4), 2351–2383. https://doi.org/10.1109/COMST.2021.3101000*

[13] *Katt, B., Beckers, K., & Wieringa, R. (2021). Cyber deception: State-of-the-art, challenges, and future directions.* Computers & Security, *108, 102376. https://doi.org/10.1016/j.cose.2021.102376*

[14] *Panda, S., Rass, S., Moschoyiannis, S., Liang, K., Loukas, G., & Panaousis, E. (2021). HoneyCar: A framework to configure honeypot vulnerabilities on the Internet of Vehicles.* IEEE Transactions on Intelligent Transportation Systems*, 23(11), 20495–20507. https://doi.org/10.1109/TITS.2021.3134201*

[15] *Iyer, K. I. (2021). Adaptive honeypots: Dynamic deception tactics in modern cyber defense.* International Journal of Scientific Research in Computer Science & Engineering, *9(1), 1–8.*

[16] *Kumar, V., Bhardwaj, S., Chouksey, P., Sadotra, P., & Chopra, M. (2021). Emerging trends in honeypot research: A review of applications and techniques.* International Journal of Human Computing Studies, *4(2), 74–88.*

[17] *Morozov, D. S., Yefimenko, A. A., Nikitchuk, T. M., Kolomiiets, R. O., & Semerikov, S. O. (2022). The sweet taste of IoT deception: An adaptive honeypot framework for design and evaluation.* Journal of Engineering and Computer Sciences*, 27(1), 122–134.*

[18] *Gopireddy, S. R. (2022). AI-powered honeypots: Enhancing deception technologies for cyber defense.* International Journal of Advanced Computer Science, *12(3), 51–58.*

[19] *Singh, A., & Joshi, R. (2022). A comprehensive review of cyber deception technologies in modern threat environments.* Journal of Cybersecurity Technology, *6(2), 178–197. https://doi.org/10.1080/23742917.2021.2005789*

[20] *Mahbooba, Z., Palomares, I., & Agiollo, Á. (2022). A multi-agent cyber deception framework for adaptive attacker engagement.* IEEE Access, *10, 119453–119468. https://doi.org/10.1109/ACCESS.2022.3219082*