Artificial Intelligence in Cybersecurity: Advancing Intelligent Threat Detection, Prevention, and Automated Response

Muhammed Sanad V. K., M. S. Bhavath Krishna, Eldhose James, Dr. Suma S.

School of Computer Science and Information Technology, Jain (Deemed to be University)

DOI: https://doi.org/10.51584/IJRIAS.2026.110400103

Received: 18 April 2026; Accepted: 24 April 2026; Published: 11 May 2026

ABSTRACT

The digital world is expanding fast, but so are the cracks in its armor. Traditional security tools are failing to keep up with the sheer volume of modern cyber threats. This paper takes a hard look at how Artificial Intelligence (AI) is stepping in to fix this mess. We're shifting from a reactive defense strategy to one that predicts attacks before they land. We explore how AI is being used to detect intrusions, classify malware, and stop phishing, while also being honest about the risks, like attackers using AI against us and the problem of opaque algorithms we can't explain. The bottom line? AI isn't a magic fix, but it's the only way we stand a chance against the speed of modern cybercrime.

INTRODUCTION

We're in the middle of a digital earthquake. The Internet of Things (IoT), cloud computing, and lightning-fast global networks have completely rewritten the rules for governments and businesses. But this connectivity is a double-edged sword: it's blown the door wide open for attackers.

Cybercrime isn't just about a lone hacker in a hoodie anymore; it's a sophisticated, multi-billion-dollar machine run by automated botnets and state-sponsored groups.

For decades, we relied on "signature-based" security. Think of it like a bouncer with a mugshot list: if you're on the list, you don't get in. That worked fine when viruses didn't change much. But today? Attackers use zero-day exploits and malware that rewrites its own code to stay invisible. The old ways are dead.

Enter Artificial Intelligence (AI) and Machine Learning (ML). Unlike the rigid rules of the past, AI learns on the job. It digests massive mountains of data to figure out what "normal" traffic looks like, so it can spot the tiny anomalies that scream "hacker." From shutting down smart phishing scams to quarantining a virus before it spreads, AI is giving us the speed we need to fight back.

METHODOLOGY

To get a clear picture of AI in the wild, we didn't just skim a few papers. We conducted a deep-dive qualitative review of the current literature. We wanted to see how different AI models—like Convolutional Neural Networks (CNNs) versus Transformers—stack up when they're put to work. The goal wasn't just to list facts, but to map out how AI has grown from a cool science project into a mandatory tool for survival.

Where We Looked

We trawled through the heavy hitters: IEEE Xplore, SpringerLink, Elsevier ScienceDirect, and Google Scholar. We focused on the years 2015 to 2025 because that's when the real magic happened—the shift from basic Machine Learning to the powerful Deep Learning we see today.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

We hunted for keywords like "AI-driven intrusion detection," "adversarial machine learning," and "automated SOAR systems." We were strict about what made the cut: if a paper was still talking about old-school signature detection, we tossed it. We only wanted research that tackles the chaotic, fast-moving threats we face right now.

How We Analyzed It

We split our findings into four battlegrounds: Intrusion Detection, Malware Analysis, Phishing Defense, and Automated Response. But we added a twist: a "Dataset Reality Check." We specifically looked at whether researchers were testing their models on nice, clean academic data (like the old KDD Cup '99 set) or if they were brave enough to use messy, real-world network logs. It turns out, models that look like geniuses in the lab often look like idiots in the real world.

Findings: AI on the Front Lines

Intrusion Detection Systems (IDS)

Traditional IDS is blind to new attacks. If it hasn't seen it before, it lets it through. That’s why researchers pivoted to anomaly detection. Early tries with things like Support Vector Machines (SVM) were okay at spotting weird math in the traffic, but they cried "wolf" way too often, as Hindy et al. 1 noted.

Now, everyone is talking about Deep Learning. Sarker 2 makes a compelling case for Convolutional Neural Networks (CNNs). Imagine treating network traffic like a picture; CNNs can "see" the shape of an attack. Recurrent Neural Networks (RNNs) are also big because they read data in a sequence, understanding the context of a connection rather than just a single moment. But Gu and Lu 3 raise a red flag: these models are "black boxes." They work, but we often have no clue *why*, which is a nightmare when you're trying to explain a breach to a CEO.

Hunting Malware

Virus writers are getting smarter, using code obfuscators to hide their tracks. So, researchers stopped reading the code and started watching what it *does*. Shelar and Rao 4 proved that deep learning can watch a program's behavior and spot malicious intent, even if the code itself looks innocent.

The cutting edge right now is hybrid models. Chaulagain et al. 5 mashed up static analysis (reading the file) with dynamic logging (watching it run). The result? A system that catches way more than either method could alone. We’re also seeing graph-based neural networks that map out malware "families," helping to catch new variants of ransomware that try to disguise themselves.

---

1 [5] Hindy, H., Brosset, D., Bayne, E., Seeam, A., Tachtatzis, C., Atkinson, R., & Bellekens, X. (2020). A Taxonomy of Network Threats and the Effect of Current Datasets on Intrusion Detection Systems. *IEEE Access*.

2 [8] Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications, and Research Directions. *SN Computer Science*.

3 [4] Gu, J., & Lu, S. (2021). An Effective Intrusion Detection Approach Using SVM with Naïve Bayes Feature Embedding. *Computers & Security*.

4 [9] Shelar, P., & Rao, S. (2021). Enhanced Capsule Network-based Executable Files Malware Detection and Classification. *Concurrency and Computation: Practice and Experience*.

5 [2] Chaulagain, B. R., Pandey, A., Basnet, S. R., & Shakya, S. (2021). Hybrid Malware Detection Using Deep Learning and Data Fusion.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

Phishing

Phishing targets people, not machines. That makes it hard to block with a firewall. Researchers are treating this as a language problem now. Karki et al. 6 used BERT—a powerful language model—to read emails and sniff out urgency and manipulation. It turns out, AI is pretty good at spotting a scammer's tone, even if the email doesn't have a virus link.

Some folks are mixing text with visuals. Sahingoz et al. 7 built a system that reads the text *and* looks at the website's layout to spot fake login pages. The problem? Scammers adapt fast. It's an arms race, and the models need constant retraining to keep up.

Automated Response

This is the holy grail: Security Orchestration, Automation, and Response (SOAR). Foley et al. 8 showed off Reinforcement Learning (RL) bots that can patch holes or quarantine infected computers on their own. But let's be real: most companies are terrified of this. If an AI makes a mistake and shuts down a critical server, heads will roll. So for now, we're sticking with "human-in-the-loop" systems.

What People Actually Think

To better understand public perception, we conducted a survey among 32 respondents, including industry professionals and students. The results show a generally positive attitude toward AI in cybersecurity, but with noticeable hesitation.

When asked whether AI improves real-time threat detection compared to traditional systems, the majority responded positively. Over half of the participants agreed, and a significant portion strongly agreed. This suggests that most respondents recognize AI’s ability to analyze large volumes of data quickly and identify threats more efficiently. However, a smaller group remained neutral, and only a few expressed disagreements, indicating that while confidence is strong, it is not unanimous.

Similarly, when questioned about AI’s ability to reduce false positive alerts in security monitoring, most participants responded favorably. A combined majority agreed or strongly agreed that AI helps minimize unnecessary alerts. That said, a notable number of respondents were neutral, and some disagreed, suggesting that concerns about accuracy and reliability still exist.

Beyond performance, trust emerged as the central issue. Nearly half of the respondents described their trust in AI as moderate. They acknowledge its usefulness but are not fully comfortable relying on it without human oversight. More concerning is that over one-quarter reported low or no trust at all. These individuals expressed worries about transparency, lack of explainability, and the risk of AI systems making decisions that are difficult to interpret or justify.

Overall, the findings indicate that while people appreciate AI’s technical strengths, full confidence has not yet been achieved. Building transparency and improving explainability will be essential for increasing trust and encouraging wider adoption.

6 [6] Karki, M., & Nasoz, F. (2022). Comparative Analysis of BERT, RoBERTa, and DistilBERT for Phishing Email Detection. *IEEE Access*.

7 [7] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine Learning Based Phishing Detection from URLs. *Expert Systems with Applications*.

8 [3] Foley, M., O'Reilly, P., & O'Sullivan, D. (2022). Autonomous Network Defence using Reinforcement Learning. *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security*.

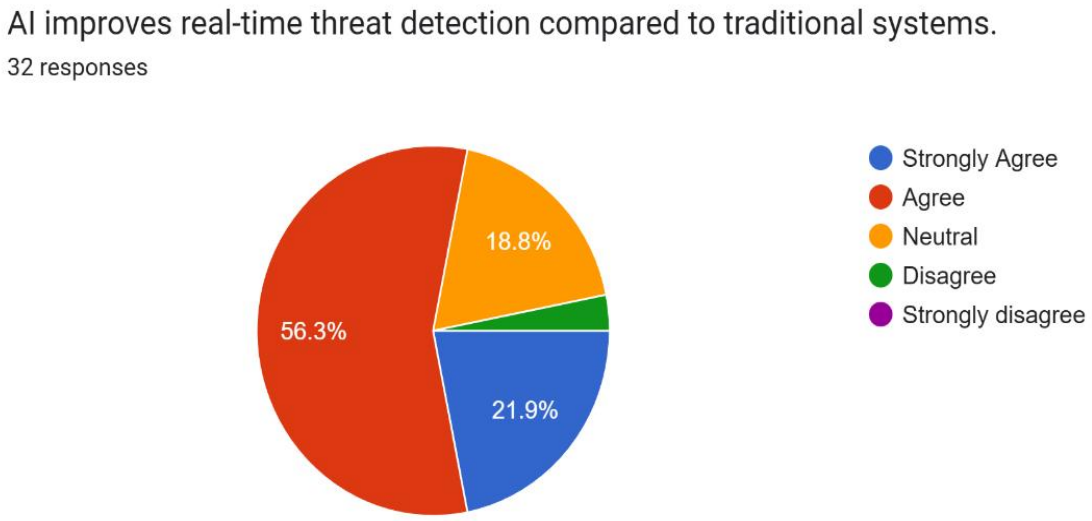

Figure1: Perception of AI Effectiveness in Real-Time Threat Detection

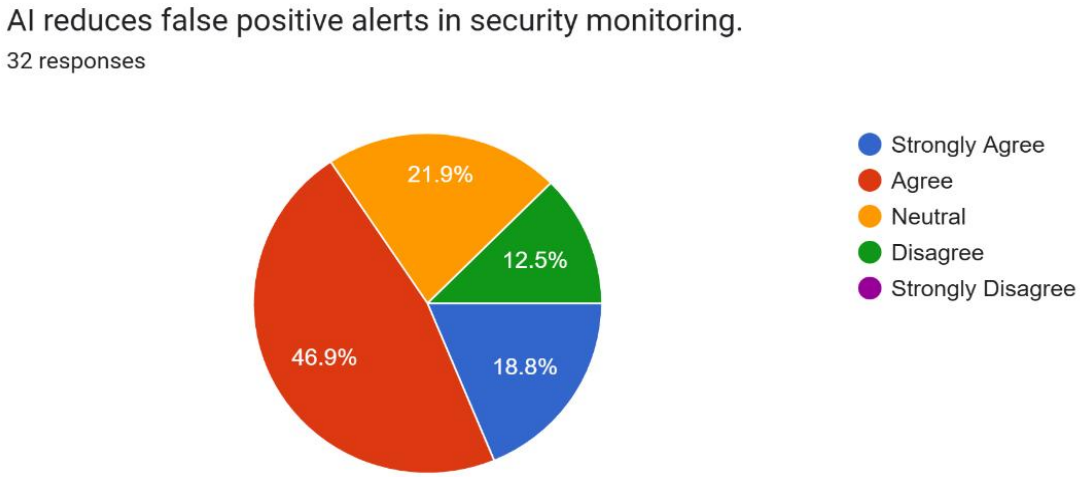

Figure 10: Perception of AI in Reducing False Positive Security Alerts

DISCUSSION:

The Elephants in the Room

The Comparison Gap

The research looks good, but there's a problem. Most of these models are trained on data that is way too clean. Old datasets like NSL-KDD don't look anything like the noise and chaos of a modern enterprise network. When you take a model trained in a sterile lab and drop it into the real world, it often falls flat. We need better, messier data.

Tunnel Vision

Most studies look at just network logs or just user behavior. Why not both? There's a huge gap in the market for systems that can cross-reference a weird network spike with the fact that "Bob from Accounting" is logging in from North Korea. Multi-modal AI is the next big thing, and we aren't there yet.

The "Black Box" and Adversarial AI

As models get smarter, they get darker. We can't see inside them. This "Black Box" issue is a dealbreaker for critical infrastructure. And it gets worse: "Adversarial AI" is real. Attackers are using their own AI to generate noise that blinds our defenses 9. If we can't explain our own tools and we can't stop them from being fooled, are we really safer? 10.

Final Thoughts

Look, integrating AI into cybersecurity isn't just a nice-to-have; it's mandatory. The threat volume is just too high for humans to handle alone. AI acts as a force multiplier, letting a small team fight back against a massive army of bots.

But let's not kid ourselves. AI isn't a silver bullet. The data sucks, the models are opaque, and the bad guys are using AI too. The future isn't about replacing humans with machines; it's about "Collaborative Intelligence." Let the AI handle the grunt work of pattern recognition, but keep a human in the pilot's seat to make the hard calls. We need systems that are tough, transparent, and trustworthy—not just high scores on a leaderboard.

REFERENCES

1. Alotaibi, A., & Rassam, M. A. (2023). Adversarial Machine Learning Attacks against Intrusion Detection Systems: A Survey on Strategies and Defense. *Future Internet*.
2. Chaulagain, B. R., Pandey, A., Basnet, S. R., & Shakya, S. (2021). Hybrid Malware Detection Using Deep Learning and Data Fusion.
3. Foley, M., O'Reilly, P., & O'Sullivan, D. (2022). Autonomous Network Defence using Reinforcement Learning. *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security*.
4. Gu, J., & Lu, S. (2021). An Effective Intrusion Detection Approach Using SVM with Naïve Bayes Feature Embedding. *Computers & Security*.
5. Hindy, H., Brosset, D., Bayne, E., Seeam, A., Tachtatzis, C., Atkinson, R., & Bellekens, X. (2020). A Taxonomy of Network Threats and the Effect of Current Datasets on Intrusion Detection Systems. *IEEE Access*.
6. Karki, M., & Nasoz, F. (2022). Comparative Analysis of BERT, RoBERTa, and DistilBERT for Phishing Email Detection. *IEEE Access*.
7. Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine Learning Based Phishing Detection from URLs. *Expert Systems with Applications*.
8. Sarker, I. H. (2021). As models get smarter, they get darker. We can't see inside them. This "Black Box" issue is a dealbreaker for critical infrastructure. And it gets worse: "Adversarial AI" is real. Attackers are using their own AI to generate.
9. . *SN Computer Science*.
10. Shelar, P., & Rao, S. (2021). Enhanced Capsule Network-based Executable Files Malware Detection and Classification. *Concurrency and Computation: Practice and Experience*.
11. Zhang, Y., & Wang, Q. (2024). Explainable Artificial Intelligence (XAI) for Cybersecurity: A Survey of Recent Trends. *Journal of Network and Computer Applications*.

9 [1] Alotaibi, A., & Rassam, M. A. (2023). Adversarial Machine Learning Attacks against Intrusion Detection Systems: A Survey on Strategies and Defense. *Future Internet*.

10 [10] Zhang, Y., & Wang, Q. (2024). Explainable Artificial Intelligence (XAI) for Cybersecurity: A Survey of Recent Trends. *Journal of Network and Computer Applications*.