Does AI Benefit Cyberattackers More? A Dynamic Game Theory Study of Ransomware Attacks in Cybersecurity

Completed Research Paper

Zhen Li
Albion College
Albion, MI, USA
zli@albion.edu

Qi Liao
Central Michigan University
Mount Pleasant, MI, USA
liao1q@cmich.edu

Abstract

Artificial Intelligence (AI) and Machine Learning (ML) are transforming the cybersecurity landscape. These technologies have shown significant promise in enhancing cyberdefense capabilities. For instance, intrusion detection system (IDS) and spam filters utilize machine learning algorithms to continuously monitor networks for abnormal behavior. However, there is increasing trend that cyberattackers are also adopting AI tools to enhance their offensive strategies. When both attackers and defenders leverage AI technologies, the balance of power may shift toward the side that can more effectively exploit AI capabilities. In this research, we study the importance of mastering AI dominance in cybersecurity contest between attackers and defenders within a dynamic game framework. Using ransomware attacks as a case study, we explore how the evolution of AI impacts the outcomes and payoffs of cyberattacks. Organizations facing cyberattack threats can utilize similar models to simulate strategic defenses and assess various levels of AI integration needed for effective cybersecurity.

Keywords: Artificial intelligence (AI), machine learning (ML), cybersecurity, adversarial/offensive AI, dynamic game theory, ransomware

Introduction

Artificial Intelligence (AI) includes machine learning (ML), deep learning (DL), natural language processing (NLP), and other computational techniques aimed at simulating human intelligence. It has been widely adopted across various industries and sectors (Rashid and Kausik, 2024; Weng, et al., 2024). As cyber threats grow in complexity and scale, advanced technologies like AI are essential to enhance the detection, prevention, and response to security incidents (Salem et al., 2024). In recent years, we have entered a new era where AI is increasingly used to assist in automating and solving many tasks for cyber defenders. Generative AI technologies like ChatGPT, Gemini, and GitHub Copilot demonstrate how AI can reduce the cognitive load and stress associated with day-to-day cybersecurity operations.

Traditionally, AI has been adopted only at the defenders' side, such as behavioral-based intrusion detection/prevention systems (IDS/IPS) using machine learning models for anomaly detections (Dong and Kotenko, 2025). As AI continues to evolve, it will enable more tailored cybersecurity solutions, helping defenders make more accurate and informed decisions. However, despite this potential, defenders must remain cautious of the possibility that attackers will also leverage AI to their advantage. Adversarial machine learning (AML), for example, has been proved effective in defeating AI/ML models used by the defenders such as spam generations (Gregory and Liao, 2023). Adversaries are increasingly turning to AI to automate their tasks, helping manage infrastructure, accelerate phishing lure creation, impersonate employees via deepfakes, and leverage open-source information and tools to develop highly tailored operational plans for threats like ransomware (Iturbe et al., 2024). AI can assist attackers in various types of cyberattack, both from a technological and a human perspective. Researchers have demonstrated that it is feasible to use generative AI to automate penetration testing via large language models (LLMs) (Gregory and Liao, 2024). By lowering the barrier to entry, AI broadens the pool of potential attackers and enhances the effectiveness of attacks by automating and scaling the attack process.

When both defenders and attackers adopt AI technologies to enhance their actions, which side will AI ultimately favor? How can we ensure that AI serves as an accelerator for cybersecurity rather than a hindrance? These are imperative research questions that warrant investigation. To that end, we formulate a novel dynamic game theoretical framework to model the interactions between cyberdefenders and cyberattackers, both empowered by AI systems. When the attack is AI-driven, targets deploy also AI-driven defenses to counter it. In game theory, a dynamic game models situations where players make decisions over time, with the order and timing of moves influencing the outcome. It captures strategic interactions where players can observe previous actions and adjust their strategies accordingly.

The AI cybersecurity dynamic game includes three key adaptations, i.e., recording of past attack success rate; adjustment of ransom requests; and AI evolution over time. The AI evolution on both the defensive and offensive sides is of special interest. We conduct extensive simulations of AI-powered cyberattacks within such dynamic game framework, using ransomware attacks as a case study. The dynamic game setting means that both the attacker and the targets continuously adjust their strategies. Each round of the game represents a full cycle of attack and defense. As the game progresses, AI on both sides evolves, making the game dynamic and non-static. In both the model and simulation, AI plays a critical role in amplifying the effectiveness of both the attack and defense.

The main contributions of this research are modeling a dual-AI adoption adversarial environment in cybersecurity and the simulation of how AI evolution can impact cybersecurity outcomes and payoffs. Our methodology involves constructing a multistage dynamic game of AI-powered ransomware attacks and defenses. We model the key stages of the ransomware lifecycle and simulate the impacts of relative AI evolution on the AI levels of the game players, the attack success rates, and the attack payoffs. The dual-AI adoption environment reflects a mutually reinforcing, co-evolving AI training dynamics between attackers and defenders.

We find that in the scenario specified by the game-theoretical setting, i.e., a game-theoretical scenario where a single attacker (or a group treated as a unified entity) launches attacks on multiple targets, when chance does not favor any player and AI evolution follows uniform rules, the attacker tends to benefit more from AI than the defenders. The repeated game with non-predefined rounds seems to be infinite as the attacker's wealth keeps accumulating, allowing the attacker to launch more attacks with the expanding budget. In more cases, the attacker ends up with a higher AI development level than the targets, probably due to the model presumption that the attacker learns from the past dealing with all potential targets while each target learns from only personal experiences dealing with the attacker. The insight derived is that although cyber-defenders may take a cutting edge in AI as a starter, AI can eventually benefit attackers more as the vast target population enables more effective training and learning for offensive AI than defensive AI. In a mutual AI learning environment, no matter how advanced the defenders' AI levels are, attackers may maintain financial incentives to launch attacks. This study highlights the importance of gaining an AI advantage in winning the cybersecurity game. Although cyberattacks such as ransomware attacks naturally give attackers a financial advantage by demanding ransoms, the relative advancement of AI by defending targets can slow down the attacker's wealth accumulation. The future balance between attackers and defenders will depend on who can innovate faster. Organizations facing cyberattack threats can use similar game models to simulate strategic defenses and evaluate the various levels of AI integration required for effective cybersecurity defense.

The rest of the paper is organized as follows. We first discuss related work. We then describe the dynamic game, including game players, strategy spaces, the multiple stages in each round of the game, and the evolution of the game. Simulation results are presented to demonstrate the impacts of AI on the profitability and the effectiveness of cyberattacks and AI evolution on both the defensive and offensive sides. The final section concludes the work and outlines directions for further research.

Related Work

There has been extensive literature exploring the capabilities and potentials of AI/ML(DL) for cyberdefense. AI-based intrusion detection systems (IDS) can detect abnormal behavior patterns and identify potential threats more effectively than traditional signature-based methods (Dong and Kotenko, 2025; Sowmya and Anita, 2023). AI-based malware detection techniques show promising results in identifying malware that is complicated and behaves in unpredictable ways (Akhtar and Feng, 2023; Bensaoud et al., 2024). AI is efficient at identifying phishing/spam by quickly detecting patterns in large volumes of data that is hard for humans to do manually (Dada et al., 2019; Xiang et al., 2011). Surveys and reviews of various AI techniques and their applications in cybersecurity disclose the AI-effectiveness in anomaly detection, threat identification, prevention against, and incident response to prevalent threats like phishing, social engineering, ransomware, and malware (Ofusori et al., 2024; Okdem and Okdem, 2024). AI-based technologies can outperform traditional approaches in organizational cybersecurity through the entire security life cycle (Jada and Mayayise, 2024). AI has achieved much success in cybersecurity solutions, and certain cybersecurity problems would only be overcome efficiently with AI (Das and Sandhane, 2021). Explorations of opportunities for further advancing AI-based cybersecurity adoptions are underway (Ferrag et al., 2025; Kaur et al., 2023).

While emerging as a powerful technology full of potentials for cybersecurity, AI has revealed also a complex landscape of cybersecurity challenges. Defense-aware adversaries adapt their strategies and techniques to circumvent AI defensive measures, posing a new set of cyberthreats (Imam and Vassilakis, 2019). The phenomena of adversarial samples inspire much research on adversarial machines learning (AML) (Costa et al., 2024; Long et al., 2022) and generative adversarial networks (GAN) (S and Durgadevi, 2021; Zhang et al., 2023b). AML has been proved effective in defeating AI/ML models used by the defenders such as spam generations (AlEroud and Karabatis, 2020; Gregory and Liao, 2023) and the manipulation of the Command and Control (C&C) channel on social media platforms (Rigaki and Garcia, 2018). Recent advances in AML demonstrate limitations and vulnerabilities of explainable AI methods (Baniecki and Biecek, 2024). It is feasible to use generative AI to automate penetration testing via large language models (LLMs) (Gregory and Liao, 2024; Iturbe et al., 2024). A survey reviews different types of attacks on AI models and the data used to train them, emphasizing the importance of developing secure and robust AI models to ensure security (Rahman et al., 2023).

In most recent years, there has been a growing trend that adversaries are increasingly turning to AI to automate their tasks, tailoring operational plans for threats like ransomware (Iturbe et al., 2024). When both adversaries and defenders use AI, it creates a dynamic and constantly evolving “arms race” where each side attempts to outsmart the other, creating mutually reinforced evolutions of both offensive and defensive AI. The dual use of AI by the opposing parties fits naturally in a dynamic game setting. Game theory is commonly used to study the complexities of cybersecurity (Dasgupta and Collins, 2019; Do et al., 2017; Ogunbodede, 2023; Verma et al., 2024).

Lastly, some researches focus on ransomware attacks, from static models of ransomware pricing (Hernandez-Castro et al., 2020), data-selling (Li and Liao, 2020), etc. to dynamic games of ransomware negotiations (Caporusso et al., 2018; Ryan et al., 2022), ransomware attack and defense (Zhang et al., 2023a; Zhao et al., 2021), etc. This paper builds a dynamic multistage game of ransomware including choosing attack targets, launching attacks and monetizing ransom, similar to researches adopting multistage game frameworks of ransomware (Ryan et al., 2022; Zhao et al., 2021). On top of that, both the ransomware attacks and defenses are AI-powered. The multistage game is repeated in non-predefined rounds, along which both the offensive and defensive AI is constantly evolving, determining the attack success rates and the corresponding payoffs of ransomware attacks. To the best of our knowledge, this paper is the first research conducting dynamic game theory and simulation studies of cybersecurity where both cyberattack and cyberdefense are AI-powered.

A Dynamic Game of AI-powered Ransomware Attack and Defense

We use ransomware as a case study to examine the implications of the simultaneous adoption of AI by both the attacker and the targets. To capture the dynamics and the interactions between the two parties, we model an iterated game between one attacker and multiple targets. The attacker is a malicious actor (or a group of malicious actors) operating within a budget (financial constraint), denoted as “B”. The targets are organizations that are targeted by the attacker for ransomware payments. All players in the game are equipped with AI, and their AI levels evolve as the game progresses.

Specifications of AI-powered Ransomware Attack and Defense

To bypass the targets' security, the attacker must develop ransomware and continually update and improve it to overcome the ever-evolving challenges of compromising the targets. This task is facilitated by AI. Automation and scaling mechanisms make attacks more sophisticated and harder to defend against. AI not only enhances the effectiveness of an attack but also fundamentally alters how attacks are executed by enabling faster replication, better adaptation, and more strategic targeting. On the defense side, the targets also leverage AI to strengthen their defenses. Defenders deploy AI-based intrusion detection systems to identify abnormal behavior patterns. Plausible countermeasures and adaptations to the attacker's AI-driven tactics include the targets' adjusting to evolving AI strategies and implementing countermeasures that can scale as quickly as the attacks themselves. This could involve AI-driven traffic filtering, dynamic network adjustments, or real-time patching of vulnerabilities.

We model AI's role in cyberattacks and cyberdefenses within a dynamic game between one attacker and N targets as follows:

- The attacker is initially equipped with AI technology to develop and improve ransomware;
- There is no additional AI training cost after the game starts, assuming the marginal cost of AI is minimal and is therefore ignored in the model;
- The AI used by the attacker is modeled to increase the attack success rate;
- The AI used by the targets is modeled to decrease the attack success rate;
- The AI levels of both the attacker and the targets evolve as the game progresses. The rate of evolution decays over time, with the decay being partially offset by the relative improvement in the AI level of the opposing party.

The game is assumed to be zero-sum, meaning the target's loss is the attacker's gain. We examine who benefits more from AI by analyzing how the relative size of the parameters impacts the outcomes of the game. The key parameters of interest are the AI levels of the attacker and the targets, as well as the corresponding ransomware attack success rate. Additionally, we explore how the relative evolution of AI affects the economic welfare of both the attacker and the targets by tracing the time path of the profitability of ransomware as AI evolves.

Structure of the Game

The attacker and the targets enter the dynamic ransomware game equipped with AI technologies, meaning the strategic actions of both parties are AI-powered. Specifically, the attacker makes an initial investment in self-learning and evasive AI ransomware, while the targets invest in predictive AI and adaptive learning.

Each round of the game is a multistage interaction between the ransomware attacker and the targets, focusing on the decision-making processes of both parties during a ransomware attack and ransom payment. In this process, we identify five stages involved in a typical t^{th} round of the game. Each round is a Stackelberg game of leaders and followers, capturing the strategic advantage the attacker has by moving first and the targets' reactive strategy.

In a new round of the game, the attacker adapts their strategy by using information from prior rounds. Both the attacker and the targets continuously update their AI capabilities in each round. The attacker starts the t^{th} round with an accumulated budget, "B^t". The attacker must decide which targets to attack and what ransom to request if a ransomware attack is successful. The attacker forms expectations about what will happen in round t based on the events that occurred in all previous rounds.

Stage 1 – Choose Targets

With AI-powered attack techniques in place, the attacker selects targets to attack based on the following rules:

- Choose targets based on expected profit, ranked from highest to lowest;
- Only select targets with expected profit; targets with expected loss are disregarded;

- Ensure that total attack costs stay within the available budget.

The budget constraint and financial rules that govern the attacker's selection of targets are as follows:

$$\sum_{n=0}^{n^t} C_n \leq B^t \tag{1}$$

and

$$\pi_n \geq 0 \tag{2}$$

where C_n is the cost of executing the attack, including AI training, infrastructure, evasion tactics, etc., n is the index of the target, and $n^t \in [0, N]$ represents the number of targets the attacker chooses to attack in the t^th round of the game. The attacker selects targets within the budget constraint by ranking them according to expected profit, i.e., $\pi_1 \geq \pi_2 \ldots \geq \pi_n \ldots \geq \pi_N$.

The adaptive attacker uses past experiences of attack successes and ransom requests to estimate the profits to be gained from individual targets as follows:

$$\pi_n = \overline{\gamma_n} \times \overline{p_n} \times R_n - C_n \tag{3}$$

where $\overline{\gamma_n}$ is the cumulative attack success rate for the target in previous rounds of the game, and $\overline{p_n}$ is the probability that the target paid ransom in the past. These two values are updated in each round. R_n is the ransom amount the attacker plans to demand from the target if the attack is successful. Hence by definition $\overline{\gamma_n} \times \overline{p_n} \times R_n$ measures the expected probabilistic ransom the attacker anticipates from target n. Given the cost of attacking the target, measured by C_n, $\overline{\gamma_n} \times \overline{p_n} \times R_n - C_n$ measures the profit the attack expects to gain from the target. The rule for demanding ransom is outlines below in Stage 3.

When $B^t = 0$, the attacker runs out of budget, and the game ends. Besides, the attacker will not launch attacks on any targets with an expected loss, even if the budget permits, meaning that $\pi_n \geq 0$ for any target that attacker may choose. In other words, the game also ends if the attacker runs out of all potentially profitable targets.

Stage 2 – Launch Attacks

The attacker launches attacks by spreading ransomware via emails, websites, or exploiting vulnerabilities. Once a target is infected, its data files are encrypted by the ransomware, and the target is then requested to pay a ransom.

The attack success rate changes as the AI adapts and improves. In each round of the game, the likelihood that a target is compromised depends on the relative AI development levels of both the attacker and the target. A relatively high AI defense level indicates stronger defenses and more effective countermeasures, while a relatively high AI attack level implies a more effective attack.

Without loss of generality, we assume that the probability of successfully infecting a target with ransomware is positively proportional to the AI level of the attacker and inversely proportional to the AI level of the target. This assumption is reasonable, as more advanced and intelligent offensive AI can learn and adapt to defender strategies quicker, identify vulnerabilities more efficiently, prioritize high-value, low-risk targets, and better evade detection. Conversely, more advanced and capable defensive AI can detect and respond to threats more rapidly, proactively strengthen defenses, and learn more effectively from past attacks. To illustrate how the attack success rate depends on the relative AI capabilities of the attacker and the target, we use the following mathematical formula:

$$\gamma_{n,t} = 1 - e^{-\frac{A_{a,t}}{A_{n,t}}} \tag{4}$$

where $\gamma_{n,t} \in [0,1]$ represents the attack success rate on target n, which is the probability that the target is successfully compromised by ransomware in round t. $A_{a,t}$ is the AI level of the attacker, and $A_{n,t}$ is the AI level of the target in the t^th round of the game, with both $A_{a,t} > 0$ and $A_{n,t} > 0$. The relative AI level of the attacker and the target is the key determinant of the attack success rate: As $\frac{A_{a,t}}{A_{n,t}} \to \infty$, $\gamma_{n,t} \to 1$; As $\frac{A_{n,t}}{A_{a,t}} \to \infty$, $\gamma_{n,t} \to 0$.

If the attack fails (i.e., $\gamma_{n,t} = 0$), the attacker receives a zero payoff and only incurs the attack costs. The game between the attacker and the n^th target pauses until the next round. However, if the attack succeeds, the game between the attacker and the target proceeds to Stage 3.

Stage 3 – Ransom Request

Upon a successful attack, the attacker requests a ransom of R_n from the n^th target, which is now a victim. The attacker's optimal ransom strategy (R_n^*) is to set the ransom just below the victim's data value:

$$R_n^* = V_n - \varepsilon \tag{5}$$

where V_n is the data value of the victim, and ε is a small amount to ensure ransom payment. In other words, the attacker aims to extract the maximum possible ransom without pushing the targets to refuse. It is a reasonable and economically grounded assumption that the optimal ransom demand is slightly below the victim's willingness to pay (WTP), which is primarily influenced by the value they assign to their data. If the attacker sets the ransom above the victim's WTP, they risk getting nothing. Setting it just below maximizes the chance of payment while extracting the highest possible value.

To request the optimal ransom, the attacker needs to know the value of the victim's data. Without relevant information, the attacker is unlikely to set the optimal ransom demand. In practice, due to information asymmetry, the attacker may not know the victim's exact WTP, but they may estimate it. In a dynamic game environment, the attacker can estimate the data value through trial and error. The attacker adopts an adaptive ransom strategy by implementing a learning mechanism, where the ransom is adjusted based on past experiences of successful or failed ransom requests. The fundamental rule is the marginal adjustment to ransom request: If the victim paid the most recent ransom demand, the attacker will raise the ransom by a margin; if the victim rejected the most recent ransom demand, the attacker will lower the ransom by a margin. In a repeated game, the marginal adjustments to the ransom request, in theory, will eventually allow the attacker to approximate the data value of the target, assuming the data value remains constant over time.

This assumption of incremental adjustment in ransom request is aligned with how some ransomware operators operate in practice. From a theoretical perspective, this kind of approach can be framed as a sequential decision-making problem or a dynamic pricing model based on observed behavior. From a strategic point of view, incremental adjustment based on behavior allows the attacker to infer the victim's WTP, especially if no prior information is available. If the victim pays the ransom, the attacker may infer the WTP is at least that amount, and possibly higher. If the victim doesn't pay, they might infer the current price exceeds WTP. Our simplified ransom adjustment strategy mirrors basic economic signaling and adaptive pricing strategies.

Stage 4 – Ransom Decision of the Victim

The victim of the ransomware attack decides whether to pay the ransom. The victim's decision variable is p_n: (i) $p_n = 1$, indicating the victim chooses to pay the ransom; and (ii) $p_n = 0$, indicating the victim chooses not to pay the ransom. The ransom payment rule is: The victim pays the ransom if the ransom demand does not exceed their WTP; otherwise, the victim refuses to pay.

To maintain focus on core strategic dynamics, the model excludes bargaining strategies and post-demand interactions, such as whether attackers return data, the completeness of recovery, or hidden exploitation of encrypted data. These factors have negligible impact on the overall conclusions. For instance, failing to return data after payment would damage attacker reputation, reduce future victim compliance, and outweigh the minor cost savings from avoiding recovery. It is therefore reasonable to assume attackers return data upon payment, allowing the data recovery stage to be omitted. Likewise, while bargaining could reveal a victim's willingness to pay, it primarily affects the ransom amount in a given instance and does not alter the study's central insights into the strategic implications of AI for attackers and defenders.

Stage 5 – AI Learning and Evolving

The AI levels of both the attacker and the targets evolve with the new data obtained in each round of the game.

AI evolution of the attacker. The rule for AI evolution of the attacker is as follows:

$$A_{a,t} = A_{a,t-1}(1 + \eta_{a,t})(1 + r_{a,t}s_{t-1}) \tag{6}$$

where $A_{a,t-1}$ is the AI level of the attacker at the end of the previous round of the game, which is also the AI level of the attacker at the beginning of the current round. $r_{a,t}$ is a random factor that reflects the randomness (and chance) in the effectiveness of AI learning and training, partially accounting for the impact of the environment on the probability of launching a successful attack. A positive and high r indicates good luck in AI training, while a negative and low r suggests that AI training is not progressing as smoothly or as expected. $s_{t-1}$is the cumulative attack success rate of the attacker in all previous rounds of the game.

In particular, $\eta_{a,t}$ is the AI learning rate of the attacker, which decreases over time to reflect diminishing marginal return to AI training, similar to the learning rate decay technique commonly used in machine learning models. As AI training progresses, the learning rate gradually decreases. In AI learning, the learning rate determines how much the AI level changes based on past experiences. Mathematical representations of learning rate decay include step decay schedules, exponential decay schedules, polynomial decay schedules, and others. Without loss of generality, we use a modified inverse time decay schedule, where the number of rounds of the game ("t") is used to reduce the learning rate through inverse decay.

In addition to time decay in AI training, the AI strength and effectiveness of an opponent can positively impact the effective training of a party. To highlight the mutually reinforcing nature of AI training in a dynamic cybersecurity game, we assume that the AI level or sophistication of the attacker can accelerate the defender's AI training, and vice versa. The idea that the attacker's AI can speed up the defender's training is rooted in adversarial machine learning, where both sides continuously adapt to one another. A more sophisticated AI attacker can introduce complex threats and new attack patterns, prompting the defender's AI to learn and mitigate these, thus accelerating its development. Similarly, an advanced AI defender may generate more complex responses, offering high-quality data — such as detection patterns, countermeasures, and behavioral shifts — for the attacker's AI to learn from. In adversarial learning, a strong opponent forces the other side to improve faster. The smarter the adversary, the more one can learn, as long as one survives long enough to observe and adapt to the evolving dynamics. In mathematical expression, we use the relative AI level of the game players as the coefficient of t. Since one can learn more effectively from a stronger opponent, an increase in the relative AI level of the opposing party will slow down the learning rate decay.

Based on the above, we use the following relative-AI-adjusted inverse time decay schedule to model the attacker's AI learning rate:

$$\eta_{a,t} = \frac{1}{1 + \frac{A_{a,t-1}}{A_{n,t-1}}t} \tag{7}$$

where ${A_{a,t-1}}/{A_{n,t-1}}$ is the AI level of the attacker relative to the nth target. The attacker's AI learning accelerates as its relative AI level decreases, which is equivalent to an increase in the relative AI level of the target.

AI evolution of the targets. While the attacker updates their AI models, the targets update their AI models as well to improve detection and response policies, as reflected by the following AI evolution rule:

$$A_{n,t} = A_{n,t-1}(1 + \eta_{n,t})(1 + r_{n,t}f_{t-1}) \tag{8}$$

where $A_{n,t-1}$ is the AI level of the nth target at the end of the previous round of the game, and $f_{t-1} = 1 - s_{t-1}$ is the cumulative rate of attack failure of the attacker, which is equivalent to the defense success rate of the target.

We use a relative-AI-adjusted inverse time decay schedule to model the AI learning rate for the target that is similar to the schedule for the attacker where ${A_{a,t-1}}/{A_{n,t-1}}$ in Equation (7) is replaced with ${A_{n,t-1}}/{A_{a,t-1}}$.

To summarize, AI-powered cyberattacks adapt to evade detection, compelling defenders to continuously retrain their AI models. As a result, both the attacker and the targets engage in AI training in each round of the game. Between rounds, the attacker updates their AI models according to their AI evolution rule, while the targets do the same based on their own evolution dynamics. In the subsequent round, the attacker ranks potential targets by expected profit, selects targets, launches attacks, and adjusts ransom demands through marginal updates informed by recent ransom outcomes. The targets, in turn, decide whether to comply with ransom requests. Through this iterative process, both sides learn from each other, continually improve their AI systems, and the game persists.

Factors Affecting AI Learning

As modeled, there are four factors that influence the progress of AI training for both the attacker and the targets.

- The duration of AI training;
- Past records of attack outcomes;
- Randomness or chance;
- The relative AI level of the attacker and the targets.

Item 1 states that AI learning slows down as the game progresses. Like regular business investment, AI training is subject to diminishing returns, meaning the AI level increases at a decreasing rate, which aligns with the concept of learning rate decay in machine learning.

Item 2 reflects the adaptive nature of learning. AI learns from past experiences. A higher cumulative attack success rate in previous rounds enhances AI learning for the attacker. Similarly, a higher cumulative attack failure rate of the attacker (i.e., a higher cumulative defense success rate of the target) enhances AI learning for the target.

Item 3 embodies the randomness or chance involved in AI training. In this context, “chance” refers to uncertainty and randomness, factors beyond direct control or prediction, that can influence the performance and effectiveness of AI training. The interactions between random initialization of parameters, data quality, hyperparameter tuning, and the stochastic nature of optimization processes can lead to unexpected outcomes, both positive and negative.

Item 4 captures the mutually-reinforced learning between the attacker and the targets. One side of the game can learn more effectively when the opponent is stronger and more sophisticated. The attacker learns faster as the relative AI level of the targets increases, and the targets learn faster as the relative AI level of the attacker increases.

Evolving Financial Conditions of the Attacker

The financial conditions of the attacker change as the game progresses. In the tth round of the game, the ransomware profit received by the attacker is:

$$\Pi^t = \sum_{n=0}^{N} (R_n - C_n) \tag{9}$$

where $R_n - C_n$ is the net revenue, or profit, received from a target. Specifically, $R_n = 0$ for targets (i) not chosen as targets, (ii) on whom the attack fails, and (iii) who reject the ransom request. Additionally, $C_n = 0$ for those not chosen as targets.

The attacker’s budget at the end of the tth (or the beginning of the (t+1)th) round of the game is

$$B^{t+1} = B^t + \Pi^t \tag{10}$$

The attacker’s budget may increase, decrease, or remain unchanged, depending on whether they earn profits, incur losses, or break even in a given round of the game.

Dynamic Game Outcomes

If the dynamic Stackelberg-type cybersecurity game is not AI-powered, the outcome is typically a Subgame Perfect Nash Equilibrium (SPNE). The attacker moves first in the game by selecting targets and setting the ransom, and the targets respond by deciding whether to pay the ransom. This sequential structure can be analyzed using backward induction, leading to SPNE. The attacker anticipates the target's best response and optimally sets the ransom just below the target's data value. The targets, in turn, rationally choose to pay the ransom if it is no higher than the value of their data. If the attacker consistently sets the ransom close to the target's data value, and the target acts rationally, the game settles into a predictable pattern.

However, with AI-powered learning and adaptation, the game may never reach a dynamic equilibrium where both players stabilize their strategies over time. In other words, the dynamic cybersecurity game between the attacker and the targets may never stabilize, and both parties will continuously evolve. The AI-powered game functions like an arms race between the attacker and the defenders, with the game's outcome depending on which party is more AI-efficient.

Without the random factors and mutual learning, expected game outcomes are as follows:

- If the attacker improves AI faster than the targets, attack success rates increase over time. Attacks become more effective, widespread, and profitable, and AI benefits the attacker more;
- If the targets improve AI faster than the attacker, attack success rates decline over time. Ransomware becomes unprofitable, and the attacker may eventually go bankrupt;
- If the AI arms race continuous without a decisive advantage, a cybersecurity equilibrium emerges where attack and defense evolve at the same rate. Both the attacker and the targets will continuously escalate tactics, and the AI arms race keeps cyber risks high for both parties.

Random factors and mutual learning introduce complexity into the game's dynamics. The time path of the game will not be unidirectional. The influence of the randomness is inherently unpredictable. The effects of mutual-learning AI on both the attacker and the targets – as well as on their interactions — depend on the efficiency, effectiveness, and reinforcement capabilities of each side's AI systems, as governed by their respective AI evolution rules. Since the game is zero-sum — meaning the victim's ransom payment (a "loss") is the attacker's revenue (a "gain") — we can trace the dynamic changes in the attacker's budget across game rounds to assess the impact of AI on both the attacker and the targets.

Simulation Study

Based on the cybersecurity attack-and-defense scenario outlined in the dynamic game, we simulate a dynamic AI-powered ransomware game between one attacker and nine targets. The definitions and assigned values of the game parameters are provided in Table 1.

Parameter	Definition	Value
N	Number of targets	9
r	Random factor in AI training	$r \in (-0.5, 0.5)$
V	Data value of the target	100 – 900 for Targets 1 to 9 with 100 incremental
C	Attack cost	20% of the data value of an individual target
A	AI level	Initial AI level is 0.5 for all
η	Attack success rate	Initial attack success rate is 50%
R	Ransom request	Initial ransom is two times the attack cost
B	Attacker's budget	Initial budget is 500

Table 1. Parameter Definitions and Assigned Values in the Simulation

The simulation runs for 100 rounds, i.e., T = 100. The attack success rate depends on the effectiveness of the attacker's AI versus the effectiveness of the targets' AI. Our developed program simulates the dynamic Stackelberg game between the attacker and multiple targets who take sequential actions. The data value of the targets ranges from 100 to 900 for Target 1 through Target 9, with an incremental increase of 100. The attack cost for each target is set to 20% of the target's data value. Both the data value and the attack cost are held constant. The attacker selects targets and requests ransom, while the targets decide whether to pay the ransom. The incremental adjustment to ransom request is 10%, i.e., ransom request is raised by 10% if the most recent ransom is paid, or ransom request is lowered by 10% if the most recent ransom request is declined. Both the attacker and the targets deploy AI in their attacks and defenses. We simulate their AI effectiveness by considering both adaptiveness and randomness.

In this dynamic game simulation, the relationship between Round *t* and Round *t+1* captures how the strategies of both the attacker and the targets evolve over time based on prior outcomes, such as adjustments to ransom demands. In particular, the successes and failures of past attacks and defenses are incorporated into the AI training processes on both sides.

Attack Profitability and Attacker Advantage

We track the attacker's total money ("budget") earned from all targets, as well as the money earned from individual targets, and monitor the AI evolution for both the attacker and the targets to gain insights into the role of AI in effective cyberattacks and cyberdefenses.

Figure 1 shows how the total money earned by the attacker changes over time. As shown, the attacker's total funds increase over time, albeit with fluctuations. This trend is unsurprising, given that ransomware attacks can be highly profitable. Since the attacker targets only those victims with expected positive returns and successfully collects ransoms from compromised systems, profits are likely to grow steadily — particularly when neither side holds a clear AI advantage. The attacker's improving financial position highlights an inherent advantage in the offensive role, with substantial potential for financial gain.

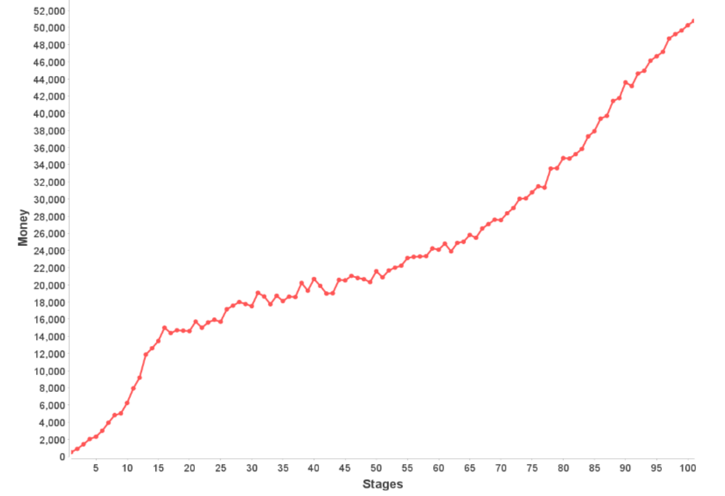

Figure 1. Attacker's Total Money

However, an improved AI defense level of the target, relative to the AI level of the attacker, is expected to reduce the profitability of ransomware attacks. More advanced defensive AI can slow the growth of attack profits and may even cause them to decrease. To illustrate the impact of the relative AI evolution between attackers and defenders on ransomware profitability, Figure 2 shows the per-target ransomware profit earned from attacking targets with varying levels of AI development. Since targets have different data values, and attacking a high-value target may yield a higher ransom in dollar terms, ransom profits are scaled by the data value of each target for better comparison.

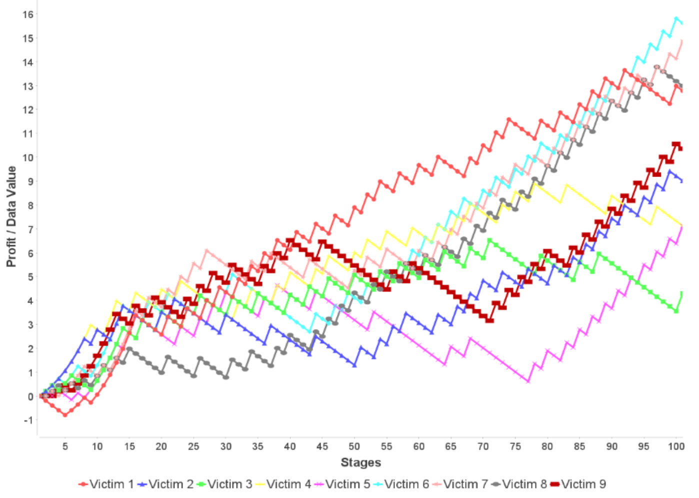

Figure 2. Data-value-adjusted Attack Profit per Victim

As shown, while ransomware profits generally increase across targets, several patterns emerge:

- Some targets demonstrate AI advantage over the attacker (Targets 3 and 4 in the simulation). When attacking these targets, the attacker experiences a slight initial increase in attack profit, followed by a decline in attack profit in later rounds of the simulation;
- Some targets start with initial AI advantage but are later overtaken by the attacker (Targets 2, 5, 7 and 9 in the simulation). When attacking these targets, the attack profit stays about the same initially, then picks up in later rounds of the simulation. There can be temporary fall in attack profit;
- The attacker dominates some targets in AI advantage throughout (Targets 6 and 8 in this simulation). When attacking these targets, attack profit increases overall.

Evolution of AI Model Levels and Attack Profitability

Figures 3-5 show the evolution of the AI levels of the attacker and targets, each representing a different pattern. The patterns of changes in attacker profit among targets are closely related to the relative AI development of the attacker and the targets. Generally, the growth of attack profit slows down when the AI level of the targets exceeds that of the attacker, and vice versa.

Figure 3 shows the evolution of the AI levels between the attacker and Target 3 representing for targets who achieve an AI advantage over the attacker as the game progresses. In this case, targets have faster AI evolution compared to the attacker in most rounds of the game, leading to a gradual increase in attack profit (with the increase in attack profit being slower than in the AI-neutral scenario). As the relative AI advantage of the targets strengthens in later rounds of the simulation, the attacker's total money decreases. Overall, the data-value-adjusted profit lines of the targets are relatively flat, as seen in Figure 2.

Figures 4 shows another representative phenomenon where AI is initially used by cybersecurity defender (e.g., training an AI model to detect attacks as an IDS). However, when attackers utilize AI as well (e.g., adversarial machine learning), they can successfully defeat the machine learning models deployed by the defenders. For Targets 2, 5, 7 and 9, we observe an initial AI advantage of the targets, which is later surpassed by the attacker. As a result, the attack profit initially increases slowly but then rises significantly, where the relative AI advantage of the attacker becomes more distinct. In the early rounds of the game, however, the AI advantage of the targets successfully keeps the attacker's funds at a low level, with limited increases from the initial budget. Target 5 is used as an example for illustration purpose.

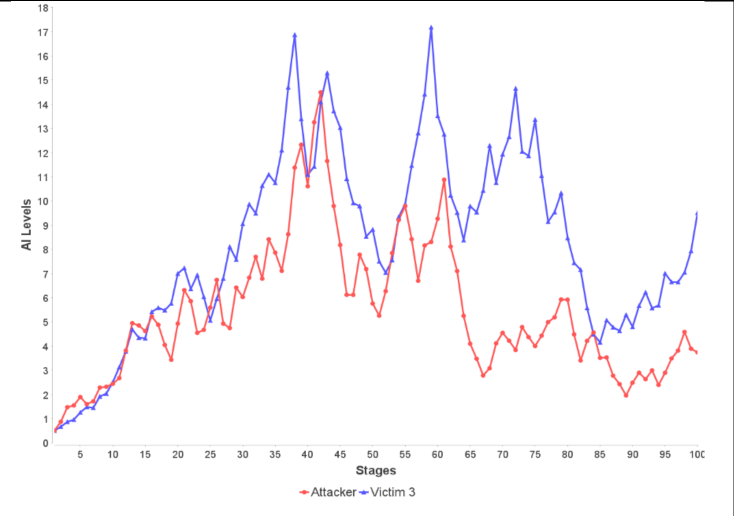

Figure 3. Game Evolution of AI Model Levels between Attacker and Defenders Who Have AI Advantage over Attacker

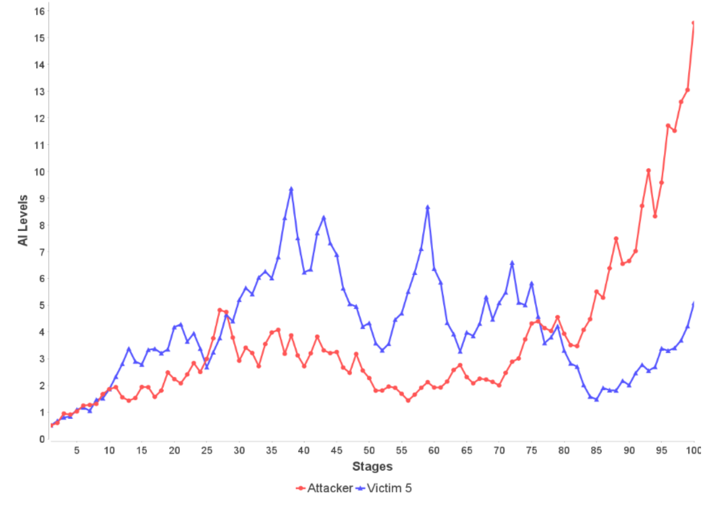

Figure 4. Game Evolution of AI Model Levels between Attacker and Defenders Who Have Initial AI Advantage but are Later Overtaken by Attacker

Figure 4 shows another representative phenomenon where AI is initially used by cybersecurity defender (e.g., training an AI model to detect attacks as an IDS). However, when attackers utilize AI as well (e.g., adversarial machine learning), they can successfully defeat the machine learning models deployed by the defenders. For Targets 2, 5, 7 and 9, we observe an initial AI advantage of the targets, which is later surpassed by the attacker. As a result, the attack profit initially increases slowly but then rises significantly, where the relative AI advantage of the attacker becomes more distinct. In the early rounds of the game, however, the AI advantage of the targets successfully keeps the attacker's funds at a low level, with limited increases from the initial budget. Target 5 is used as an example for illustration purpose.

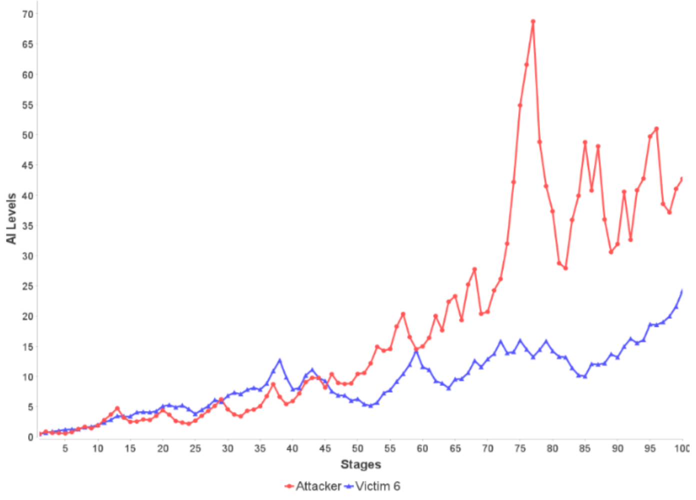

Figure 5. Game Evolution of AI Model Levels between Attacker and Defenders Where Attacker Has Dominant AI Advantage over Defenders

Finally, Figure 5 shows significant AI progress for the attacker compared to the targets (6 and 8), resulting in a faster accumulation of attack profit. Target 6 is used as an example to illustrate this pattern.

Evolution of AI Learning

In the context of the model setting, the key determinant of the relative AI evolution is the AI learning rate (η), which decays over time but increases when the opponent is AI-strong. Figure 6 compares the AI levels of Target 1 and the attacker, and Figure 7 compares their respective AI learning rates.

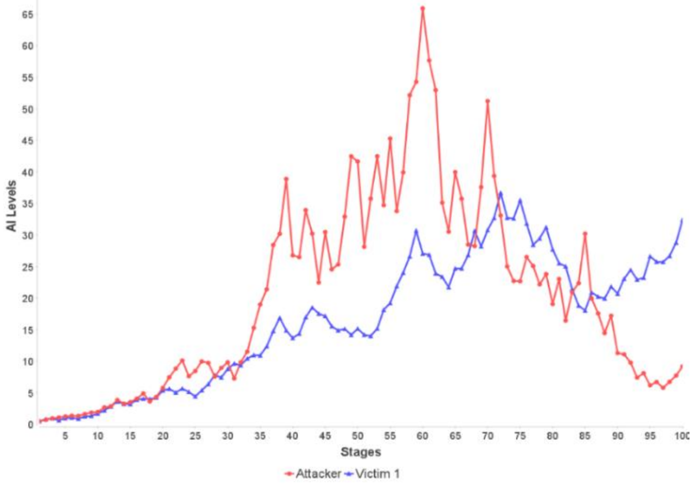

Figure 6. Game Evolution of Relative AI Levels for Attacker and Target 1

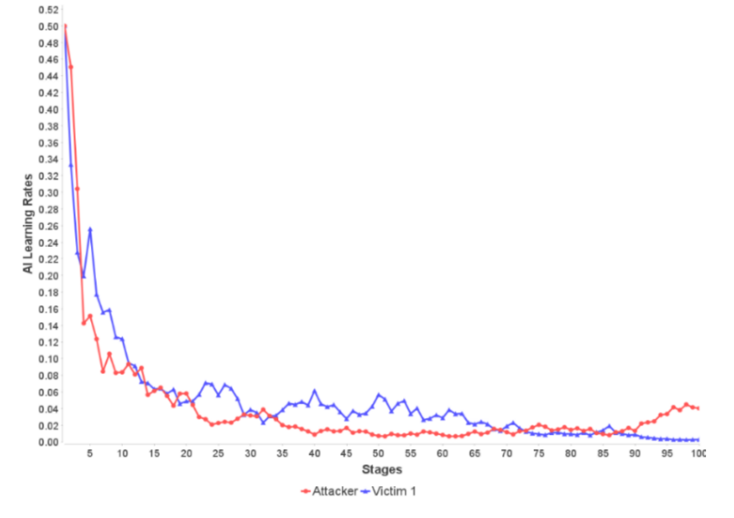

Figure 7. Game Evolution of Relative AI Learning Rates for Attacker and Target 1

We can observe that the relative AI learning rates are inversely related to the relative AI levels. When one party's relative AI level strengthens, it promotes the AI learning of the opponent. Additionally, Target 1 ranks the highest in terms of profitability for the attacker in the middle rounds of the simulation due to the significant AI advantage of the attacker, which is unique among the targets. The attacker's AI strength stimulates Target 1 to improve its AI learning. However, as the victim gradually surpasses the attacker in AI development, the attack profit decreases.

Further Discussions

Randomness in AI training introduces minor variations across simulation runs. The results presented are from a representative run, with robustness verified through multiple sensitivity analyses. While numerical values differ, all simulations consistently show: (1) increasing attack profitability, (2) distinct patterns in the relative AI evolution of attackers and targets, and (3) an inverse relationship between attack profitability growth and target AI development. These recurring patterns confirm the robustness of the study's insights.

Although research on adversarial game theory in AI-enabled cyber offense and defense remains nascent and comprehensive datasets are limited, recent trends and reports provide partial empirical support for the simulation findings.

First, AI-powered attacks present substantial financial incentives for adversaries, with illicit gains showing a persistent upward trend. In 2024, cryptocurrency scam revenues totaled at least $9.9 billion, and possibly up to $12.4 billion, driven largely by AI-enhanced attack methods (Mattackal, 2025). Ransomware remains a dominant form of crypto-related crime, generating $1.25 billion in 2023, the highest on record. Although revenues fell in 2024 following law enforcement actions against major groups such as LockBit and BlackCat, the number of ransomware incidents reported on dark-web leak sites reached an all-time high (Chainalysis, 2025).

Second, the net advantage of AI depends on how effectively each side adapts to AI-driven innovations. Between March 2024 and February 2025, one in six breaches involved AI-enabled attacks (IBM, 2025). In 2025, 68% of cybersecurity analysts reported AI-generated phishing as harder to detect than ever, while reported AI-enabled cyberattacks rose by 47% (Namase, 2025). On defense, generative AI has accelerated phishing response preparation from hours to minutes. Organizations with extensive AI and automation use saved an average of $1.9 million per breach and reduced detection and containment times by 80 days (IBM, 2025). Reflecting this reliance, the global AI-in-cybersecurity market is projected to grow from $15 billion in 2021 to $135 billion by 2030 (Acumen Research and Consulting, 2022).

Finally, long-term dynamics may favor attackers due to richer opportunities for training and adaptation in a large, diverse target population. Potential targets vastly outnumber attackers, and offensive actors tend to share tools, tactics, and intelligence more freely than defenders. Ransomware-as-a-Service (RaaS) ecosystems exemplify this, enabling rapid dissemination of capabilities, such as in the REvil group, where a core team developed the malware and affiliates distributed it. By contrast, defenders often restrict breach data sharing for privacy, reputational, or legal reasons, limiting collective learning. Many organizations lack visibility into broader attack patterns until after major incidents, reinforcing the asymmetry observed in simulation models (IBM, 2025). These factors suggest that in a sustained AI arms race, attackers' broader and faster learning potential may yield a strategic long-term advantage.

Conclusion

AI plays a crucial role in enhancing cybersecurity defenses by improving the efficiency, accuracy, and speed of detecting, preventing, and responding to cyber threats. However, as a double-edged sword, AI can assist both attackers and defenders. It can lower the threshold for launching cyberattacks while simultaneously enhancing their effectiveness. AI has demonstrated its ability to identify system and network vulnerabilities more efficiently than traditional methods, reducing the time and skill needed for attackers to compromise systems. Additionally, AI tools can automate the crafting of realistic phishing messages and malicious websites. By learning from interactions, AI can adapt its tactics to target specific individuals with high precision, thereby increasing the likelihood of successful attacks. Thus, the growing use of AI presents both opportunities and challenges for cybersecurity practices.

When AI systems are employed for both defensive and offensive purposes, the question of which side AI favors becomes central. This dynamic creates both fascinating and challenging interactions. In a dynamic cybersecurity game where both attackers and defenders utilize AI and continually learn from each other, the outcomes can be complex.

To explore this critical research question, we modeled AI-powered cyberattacks and cyberdefenses within a dynamic game theoretic framework, using ransomware attacks/defenses as a case study. Each round of the game represents a multistage process that captures the full lifecycle of ransomware attacks. We simulated the evolution of this dynamic game, tracking changes in the attacker's financial condition, as well as the AI adaptation and evolution of both attackers and targets. Special attention was given to the role of mastering AI dominance in determining the outcome of the cybersecurity contest between attackers and defenders.

Using ransomware as a case study, our simulation results indicate that, in most cases, the attacker's financial condition continues to improve, providing strong financial incentives to keep launching attacks. These results also suggest that AI is more likely to benefit attackers, especially as attackers gain AI dominance in more cases. This is likely due to the game structure, where the attacker learns from interactions with all targets, while each target learns only from interactions with the single attacker.

We examined how the evolution of AI influences the outcomes and payoffs of cyberattacks. Our findings show that whether AI benefits attackers or defenders more depends largely on how adaptive each party is to AI advancements. If AI-powered defenses improve detection and proactively patching faster than the attacker's AI innovations, then AI benefits the targets (shown by the decelerated wealth accumulation of the attacker). Conversely, if AI enhances the adaptability and automation of attacks more than it boosts defenses, AI is more advantageous for the attacker (shown by the accelerated wealth accumulation of the attacker), a trend likely to increase in the future.

The key implication is that, even if defenders initially have the advantage with cutting-edge AI, attackers can eventually gain more value from AI due to the larger target population, which enables more effective training and learning for offensive AI than defensive AI. In a mutual AI learning environment, no matter how advanced the defenders' AI may be, attackers will always have financial incentives to continue launching attacks. Organizations facing cyberattack threats can adapt the proposed models to simulate strategic defenses against rapidly evolving AI-enabled attacks. Effective cybersecurity management requires not only integrating AI but also establishing continuous learning loops that leverage real-time threat intelligence and incident data, ensuring both competitiveness and resilience.

References

Acumen Research and Consulting. (2022). Artificial intelligence in cybersecurity market analysis – global industry size, share, trends and forecast 2022-2030 (Report ID: ARC3019). https://www.acumenresearchandconsulting.com/artificial-intelligence-in-cybersecurity-market

Akhtar., M. S., & Feng, T. (2023). Evaluation on machine learning algorithms for malware detection. *Sensors*, 23(2), 946. https://doi.org/10.3390/s23020946

AlEroud, A., & Karabatis, G. (2020). Bypassing detection of URL-based phishing attacks using generative adversarial deep neural networks. *Proceedings of the 6th International Workshop on Security and Privacy Analytics*, 53-60. https://doi.org/10.1145/3375708.3380315

Baniecki, H., & Biecek, P. (2024). Adversarial attacks and defenses in explainable artificial intelligence: A survey. *Information Fusion*, 107, 102303. https://doi.org/10.1016/j.inffus.2024.102303

Bensaoud, A., Kalita, & J., Bensaoud, M. (2024). A survey of malware detection using deep learning. *Machine Learning with Applications*, 16, 100546. https://doi.org/10.1016/j.mlwa.2024.100546

Caporusso, N., Chea, S., & Abukhaled, R. (2019). A game-theoretical model of ransomware. *Proceedings of International Conference on Applied Human Factors and Ergonomics (AHFE)*, 69-78. https://doi.org/10.1007/978-3-319-94782-2_7

Chainalysis. (2025). The 2025 crypto crime report. https://go.chainalysis.com/2025-Crypto-Crime-Report.html

Costa, J. C., Roxo, T., Proença, H., & Inácio, P. R. M. (2024). How deep learning sees the world: A survey on adversarial attacks & defenses. *IEEE Access*, 12, 61113-61136. https://doi.org/10.1109/ACCESS.2024.3395118

Dada, E. G., Bassi., J. S., Chiroma, H., Abdulhamid, S. M., Adetunmbi, A. O., & Ajibuwa, O.E. (2019). Machine learning for email spam filtering: review, approaches and open research problems. *Heliyon*, 5(6), e01802. https://doi.org/10.1016/j.heliyon.2019.e01802

Das, R., & Sandhane, R. (2021). Artificial intelligence in cyber security. *Journal of Physics: Conference Series*, 1964, 042072. https://doi.org/10.1088/1742-6596/1964/4/042072

Dasgupta, P., & Collins, J. B. (2019). A survey of game theoretic approaches for adversarial machine learning in cybersecurity tasks. *AI Magazine*, 40(2), 31-43. https://doi.org/10.1609/aimag.v40i2.2847

Do, C. T., Tran, N. H., Hong, C., Kamhoua, C. A., Kwiat, K. A., Blasch, E., Ren, S., Pissinou, N, & Iyengar, S. S. (2017). Game Theory for Cyber Security and Privacy. *ACM Computing Surveys (CSUR)*, 50(2), 1-37. https://doi.org/10.1145/3057268

Dong, H., & Kotenko, I. (2025). Cybersecurity in the AI era: Analyzing the impact of machine learning on intrusion detection. *Knowledge and Information Systems*. https://doi.org/10.1007/s10115-025-02366-w

Ferrag, M. A., Alwahedi, F., Battah, A., Cherif, B., Mechri, A., Tihanyi, N., Bisztray, T., & Debbah, M. (2025). Generative AI in cybersecurity: A comprehensive review of LLM applications and vulnerabilities. *Internet of Things and Cyber-Physical Systems*, 5, 1-46. https://doi.org/10.1016/j.iotcps.2025.01.001

Gregory, J., & Liao, Q. (2023). Adversarial spam generation using adaptive gradient-based word embedding perturbations. *Proceedings of IEEE International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)*, 1-5. https://doi.org/10.1109/AIBThings58340.2023.10292495

Gregory, J., & Liao, Q. (2024). Autonomous cyberattack with security-augmented generative artificial intelligence. *Proceedings of IEEE International Conference on Cyber Security and Resilience (CSR)*, 270-275. https://doi.org/10.1109/CSR61664.2024.10679470

Hernandez-Castro, J., Cartwright, A., & Cartwright, E. (2020). An economic analysis of ransomware and its welfare consequences. *Royal Society Open Science*, 7(3). https://doi.org/10.1098/rsos.190023

IBM. (2025). Cost of a data breach report 2025: the AI oversight gap. https://www.ibm.com/reports/data-breach

Imam., N., & Vassilakis, V. G. (2019). A survey of attacks against Twitter spam detectors in an adversarial environment. *Robotics*, 8(3), 50. https://doi.org/10.3390/robotics8030050

Iturbe, E., Llorente-Vazquez, O., Rego, A., Rios, E., & Toledo, N. (2024). Unleashing offensive artificial intelligence: Automated attack technique code generation. *Computers & Security*, 147, 104077. https://doi.org/10.1016/j.cose.2024.104077

Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, 8(2), 100063. https://doi.org/10.1016/j.dim.2023.100063

Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion*, 97. https://doi.org/10.1016/j.inffus.2023.101804

Li, Z., & Liao, Q. (2020). Ransomware 2.0: to sell, or not to sell a game-theoretical model of data-selling ransomware. *Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES)*, 1-9. https://doi.org/10.1145/3407023.3409196

Long, T. Gao, Q., Xu, L., & Zhou, Z. (2022). A survey on adversarial attacks in computer vision: Taxonomy, visualization and future directions. *Computers & Security*, 121, 102847. https://doi.org/10.1016/j.cose.2022.102847

Mattackal, L. P. (2025, February 14). Crypto scams likely set new record in 2024 helped by AI, Chainalysis says. Reuters. https://www.reuters.com/technology/crypto-scams-likely-set-new-record-2024-helped-by-ai-chainalysis-says-2025-02-14/

Namase, R. (2025, July 22). AI cyber attacks statistics 2025: how attacks, deepfakes & ransomware have escalated. *SQ Magazine*. https://sqmagazine.co.uk/ai-cyber-attacks-statistics/

Ofusori, L., Bokaba, T., & Mhlongo, S. (2024). Artificial intelligence in cybersecurity: A comprehensive review and future direction. *Applied Artificial Intelligence*, 38(1). https://doi.org/10.1080/08839514.2024.2439609

Ogunbodede, O. O. (2023). Game theory classification in cybersecurity: A survey. *Applied and Computational Engineering*, 2(1), 670-678. https://doi.org/10.54254/2755-2721/2/20220644

Okdem, S., & Okdem, S. (2024). Artificial intelligence in cybersecurity: A review and a case study. *Applied Sciences*, 14(22), 10487. https://doi.org/10.3390/app142210487

Rahman, M. M., Siddika Arshi, A., Hasan, M. M., Farzana Mishu, S., Shahriar, H., & Wu, F. (2023). Security risk and attacks in AI: A survey of security and privacy. *Proceedings of IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, 1834-1839. https://doi.org/10.1109/COMPSAC57700.2023.00284

Rashid, A. B., & Kausik, M. A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, 7. https://doi.org/10.1016/j.hybadv.2024.100277

Rigaki, M., & Garcia, S. (2018). Bringing a GAN to a knife-fight: Adapting malware communication to avoid detection. *Proceedings of 2018 IEEE Security and Privacy Workshops (SPW)*, 70-75. https://doi.org/10.1109/SPW.2018.00019

Ryan, P., Fokker, J., Healy, S., & Amann, A. (2022). Dynamics of targeted ransomware negotiation. *IEEE Access*, 10, 32836–32844. https://doi.org/10.1109/ACCESS.2022.3160748

S, K., & Durgadevi, M. (2021). Generative adversarial network (GAN): A general review on different variants of GAN and applications. *Proceedings of the 6th International Conference on Communication and Electronics Systems (ICCES)*, 1-8. https://doi.org/10.1109/ICCES51350.2021.9489160

Salem, A. H., Azzam, S. M., Emam, O.E., & Abohany, A. A. (2024). Advancing cybersecurity: A comprehensive review of AI-driven detection techniques. *Journal of Big Data*, 11, 105. https://doi.org/10.1186/s40537-024-00957-y

Sowmya, T., & Mary Anita, E. A. (2023). A comprehensive review of AI based intrusion detection system. *Measurement: Sensors*, 28, 100827. https://doi.org/10.1016/j.measen.2023.100827

Verma, R., Koul, S., Ajaygopal, K. V., & Singh, S. (2024). Exploring game theoretic applications in cyber security. *Proceedings of International Conference on Intelligent Systems for Cybersecurity (ISCS)*, 1-6. https://doi.org/10.1109/ISCS61804.2024.10581244

Weng, Y., Wu, J., Kelly, T., & Johnson, W. (2024). Comprehensive overview of artificial intelligence applications in modern industries. https://doi.org/10.48550/arXiv.2409.13059

Xiang, G., Hong, J., Rose, C. P., & Cranor, L. (2011). CANTINA+: A feature-rich machine learning framework for detecting phishing web sites. *ACM Transactions on Information and System Security (TISSEC)*, 14(2), 1-28. https://doi.org/10.1145/2019599.2019606

Zhang, C., Luo, F., & Ranzi, G. (2023a). Multistage game theoretical approach for ransomware attack and defense. *Proceedings of IEEE Transactions on Services Computing*, 16(4), 2800-2811. https://doi.org/10.1109/TSC.2022.3220736

Zhang, C., Yu, S., Tian, Z., & Yu, J. J. Q. (2023b). Generative adversarial networks: A survey on attack and defense perspective. *ACM Computing Surveys*, 56(4), 1-35. https://doi.org/10.1145/3615336

Zhao, Y., Ge, Y., & Zhu, Q. (2021). Combating ransomware in internet of things: A games-in-games approach for cross-layer cyber defense and security investment. *Proceedings of the 12th International Conference on Decision and Game Theory for Security (GameSec)*, 208-228. https://doi.org/10.1007/978-3-030-90370-1_12