AI-Driven Cyber Deception in FinTech: An Adaptive Defense Strategy

Isaac Ojeh[1], Xavier Palmer[2] and Lucas Potter[2]
[1]MorphHats InfoSecure, Waterloo, Canada
[2]BiosView Labs, Dayton, Ohio

isaac.ojeh@gmail.com
Biosview1@proton.me

Abstract: FinTech platforms clear high-value transactions in milliseconds, making them lucrative targets for adversaries who increasingly weaponize artificial intelligence. Once an attacker bypasses the perimeter, via credential stuffing, supply-chain malware, or deep-fake social engineering, traditional defenses often alert too late to prevent loss. We present an Adaptive Deception Defense Framework (ADDF) that intertwines AI-orchestrated honeypots, honeytokens and decoy micro-services within everyday banking and payment workflows. A recurrent-neural threat profiler classifies live attacker behavior; a Proximal-Policy-Optimization agent then selects actions such as spawning a shadow login API, cloning a database or injecting synthetic ledgers, thereby misdirecting intruders while harvesting telemetry. In a controlled "FinBank" test-bed featuring a vulnerable Flask-and-MySQL stack, ADDF shortened mean time-to-detect from 3 min 42 s to 29 s, increased attacker dwell-time inside decoys to 12 min 18 s and prevented all real data exfiltration across ten attack trials. False-positive alerts remained below 1% per run and added resource use averaged 14% CPU/RAM on mid-range servers. The framework also produced high-fidelity indicators of compromise, password lists, malware binaries and lateral-movement scripts, that would have been unavailable under baseline controls. These findings indicate that AI-driven cyber deception can transform FinTech security from passive monitoring into proactive engagement, mitigating breach impact while supplying rich threat intelligence. The paper details system architecture, reinforcement-learning policy training, empirical evaluation and operational implications—showing how defenders can regain initiative in the AI-to-AI cyber arms race without disrupting legitimate customers or breaching regulatory duties. FinTech platforms process high-value transactions at internet speed, making them prime targets for advanced cyber-criminals who now weaponize artificial intelligence. Traditional controls detect many incidents yet remain reactive; once adversaries bypass the perimeter, defenders struggle to contain damage fast enough to prevent data loss or fraud. We present an AI-driven cyber-deception framework that inserts a dynamic layer of honeypots, honeytokens and decoy services into a live FinTech environment. A learning engine classifies attacker behaviour in real time, then deploys or adapts decoys to misdirect adversaries while capturing rich telemetry. In a controlled banking testbed, the system cut mean time-to-detect from minutes to seconds, confined intruders to fake assets in every trial, and prevented exfiltration of real customer data. Adaptive deception also generated high-quality threat intelligence with negligible false positives and modest resource overhead (<15% CPU/RAM on a mid-range server). Findings from this study suggest that the use of AI-powered deception methods can shift or otherwise redirect Fintech defence posture from passive monitoring configurations to that of proactive engagement. This can reduce risk for defensive teams and potentially boost incident-response agility without significantly disrupting legitimate users within a system where the novelty lies in orchestrating established AI components (RL, anomaly classification, and generative decoys) into a closed-loop deception system for real-time FinTech operations.

Keywords: AI-riven, Deception, ADDF, Honeypots, Machine learning, Adaptive, Proactive, Defense, Fintech

1. Introduction

Digital finance now moves trillions of US dollars annually through mobile wallets, real-time gross-settlement rails and open-banking APIs (World Bank 2025). The same connectivity and agility that enable more convenient customer interactions create ever-wider attack surfaces. In 2024 the financial sector recorded a 53% year-on-year rise in confirmed breaches, driven largely by credential-stuffing bots, ransomware campaigns against payment processors and AI-assisted social-engineering scams (Bonini 2025). FinTech firms face asymmetric risk. Adversaries need only one foothold (phished credentials, a misconfigured S3 bucket or a single zero-day) to pivot laterally and exfiltrate sensitive data. Regulators such as the European Union and the Payment Card Industry fine institutions heavily for breaches and mandate near-real-time incident response (European Union 2016; PCI Security Standards Council 2024). Meanwhile, defenders wrestle with alert fatigue: a modern Security Operations Centre (SOC) may review millions of daily events but still miss the handful that matter (Crandall 2025). Standard safeguards (firewalls, multi-factor authentication, fraud-scoring models) remain essential but inherently reactive. Once an attacker phishes valid credentials or finds a zero-day vulnerability, lateral movement can occur in seconds, threatening funds, personally identifiable information and market reputation. Operators need mitigations that buy time, mislead adversaries and reveal their tools of trade before damage occurs.Cyber deception answers that need. By scattering decoys indistinguishable from production systems, defenders turn the network into a minefield: attackers waste effort on fake assets, while defenders observe every move. Early honeypots proved the concept but suffered from static content and high upkeep. Recent

advances in artificial intelligence (especially reinforcement learning (RL) and generative models) enable adaptive deception: decoys that spawn, mutate or retire in response to live attacker cues. Cyber deception offers a complementary strategy. Defenders can use cyber deception to frustrate and create a hostile environment for intruders through the use of realistic decoys, in the form of honey pots, honey tokens, fake micro-services, and more, within production networks. Hints of decoys alone are inherently suspicious, and these can produce a high-fidelity signal that cuts through logging noise (Ferguson-Walter, 2025). Early honeypots, however, were static: attackers could fingerprint stale banners, implausible directory trees or isolated IP subnets (Moric et al. 2025). Artificial intelligence applied to defensive cybersecurity maneuvers, especially through the application its subset, machine learning, changes the equation of defense strategy outcomes. Machine learning, itself, can be utilized to high success in improving defense through uses in (i) *detecting anomalies* in high-volume telemetry (Nobre et al. 2023); (ii) *optimizing decisions* under uncertainty via reinforcement learning (Ngo et al. 2023); and (iii) *generating synthetic artefacts* that pass human or machine scrutiny (Ahmed et al. 2025). Combining these capacities with deception yields adaptive honeynets that spawn, mutate or retire in real time, mirroring genuine customer activity and data schemas. Attackers are lured into a responsive "hall of mirrors," wasting effort while defenders observe tactics, techniques and procedures (TTPs) in a safe sandbox.

This paper makes four contributions:

1. **An ADDF Architecture,** an end-to-end AI-driven deception framework purpose-built for FinTech micro-service meshes is presented
2. **A Reinforcement-Learning Policy** wherein a PPO agent trained to maximize attacker engagement and minimize risk to genuine assets is given.

- *The authors recognize that PPO is not a new concept; rather, we encourage the reader to view this work as an innovative integration of established AI techniques that can be applied within critical contexts. This can be applied to any industry that seeks to utilize fintech as part of their models; this is itself relevant to Industry 4.0 (Ferraro et al, 2024).*
- *However, it can be an orchestration policy that integrates reinforcement learning with real-time anomaly classification and generative decoy synthesis for FinTech micro-services. We use PPO as a robust optimizer (not as a novelty in itself) to decide when and where to deploy, mutate, or retire decoys under live resource, latency, and regulatory constraints. The contribution is the design and integration of this control loop in a high-stakes financial setting, not the PPO algorithm per se.*
- *An RL-driven orchestration policy (implemented with PPO as a stable optimizer) that fuses anomaly classification and generative decoying to make real-time placement decisions under FinTech resource and regulatory constraints (our novelty lies in the integration and operational design, not in PPO itself).*

3. **Generative Data Pipelines**, tools that create synthetic customer records, KY-C documents and transaction streams indistinguishable from production baselines, are demonstrated.
4. **Empirical Validation** through controlled experiments showing that ADDF cuts detection latency an order of magnitude and blocks breach objectives at modest cost is given.

Section 2 surveys threat trends and related research. Section 3 details the architecture, RL training and synthetic-data generation. Section 4 outlines our banking test-bed and attack scenarios. Section 5 reports quantitative and qualitative results. Section 6 discusses trade-offs, limitations and comparison with prior art. Section 7 explores operational and regulatory implications for FinTech. Section 8 concludes and sketches future extensions, including large-language-model (LLM) decoy chatbots and cost-aware cloud auto-scaling.

2. Background and Related Work

2.1 FinTech Threat Landscape

Typical FinTech stacks comprise web front-ends, RESTful gateways, cloud databases, blockchain nodes and vendor APIs. Prominent attack avenues include:

- **Credential abuse:** Botnets replay leaked username–password pairs against login APIs, bypassing multi-factor where SMS intercept or session riding is viable.
- **Data-layer exploits:** SQL injection, insecure deserialization or Server-Side Request Forgery against micro-services expose customer PII and payment tokens.
- **Insider privilege misuse:** Contractors or rogue employees export customer lists or abuse admin APIs.
- **AI-enabled fraud:** Deep-fake video calls trick customer-service reps; generative-AI phishing emails lift brand tone and user context (RoX818 2025).

Regulations amplify consequences. Under GDPR a cross-border breach of EU resident data risks fines up to 4% of global turnover (European Union 2024). PCI Security Standards Council v4.0 requires near-immediate detection of card-holder-data exposure and zero tolerance for unencrypted storage (PCI SSC 2024).

High velocity and regulatory penalties demand controls that stop fraud before funds leave the building.

2.2 Evolution of Cyber Deception

Honeypots as commonly known to cybersecurity practitioners date back to the 1990s; attackers quickly learned to fingerprint their static banners (Cheswick, 1992; Acal et al, 2015). Honeytokens (fake credentials/records) scale better but lack interaction depth. Modern deception platforms automate thousands of decoys that blend into production networks (Crandall 2025). Studies show decoys attract 80+% of exploit traffic when asset ratios are below 20% (Ferguson-Walter 2023). The seminal *honeypot* concept (Spitzner 1999) placed a sacrificial server to observe exploits. High-interaction honeypots emulate full operating system (OS) stacks; low-interaction variants simulate protocols cheaply. Honeytokens (Cheswick 2003) expanded deception to static artefacts (fake SSH keys, bogus credit-card numbers) embedded in live systems; any use of such tokens signals compromise. Modern deception platforms instrument thousands of virtual assets, scaling ratio such that <25% decoys attract >80% exploit traffic (Ferguson-Walter 2025). Yet static decoys suffer recognition: veteran adversaries scan for tell-tale process lists, improbable uptime or isolated subnets (Moric et al. 2025). Dynamic deception emerged: game-theoretic placement optimizes where to seed traps; moving-target defense randomizes IP addresses; but manual upkeep remains heavy.

2.3 AI for Adaptive Deception

AI can bolster deployment of deception in defensive strategies in three dimensions:

- Detection: Unsupervised models can be deployed to flag deviations from baseline behavior with a higher recall compared to signature intrusion detection systems (IDS) (Nobre et al. 2023).
- Decision-making: RL agents can learn optimal decoy allocations under resource constraints, outperforming heuristic schedules in some contexts (Ngo et al. 2023).
- Generation: LLMs produce contextually correct faux artefacts (synthetic audit logs, phishing lures or chat dialogue) improving trap realism (Ahmed et al. 2025).

Crandall (2025) describes an enterprise platform where AI mirrors production topologies hourly, ensuring traps share hostnames, certificates and load profiles with genuine servers. The SPADE framework achieved 93% attacker engagement by allowing an LLM to craft tailored ploys mid-session (Ahmed et al. 2025). Our work differs by integrating anomaly detection, RL placement and generative content specifically for FinTech workloads, then validating the combined effect empirically.

In other words, AI enhances deception through:

- **Classification**: Detecting malicious sequences in logs.
- **Reinforcement learning**: Choosing decoy placement/actions under uncertainty.
- **Generative engines**: Producing synthetic artefacts (e.g., fake ledgers) that match live schemas.

Ahmed et al. (2025) used large language models to autogenerate phishing-lure emails and achieved 93% engagement. Our work focuses on FinTech-specific services (payments, accounts) and evaluates a full adaptive loop, from detection to autonomous decoy orchestration.

2.4 Human Factors and Cognitive Bias

Attackers exhibit cognitive shortcuts such as confirmation bias and overconfidence once an initial credential proves valid. Johnson et al. (2021) found that red-teamers often assume authenticity if three successive commands succeed, ignoring subtle anomalies. Exploiting such biases, adaptive deception nudges adversaries deeper into decoys via timed success/failure cues

3. System Architecture and Methodology

3.1 High-Level Architecture

Monitoring & Analytics nodes (Zeek on network taps, OSQuery on hosts, log shipper on API gateway) stream events to a Kafka bus. Raw streams feed an **LSTM Anomaly Detector** tuned to sessions, flows and SQL patterns. Detected anomalies carry context: source IP, request type, confidence.

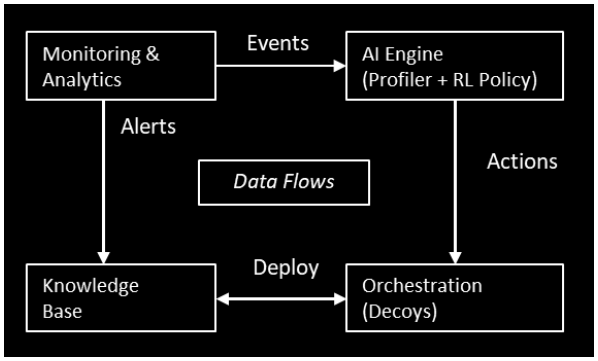

Figure 1: Data Flows

The **AI Engine** executes two models:

- **Threat-Stage Classifier** — an LSTM–Softmax stack categorizing activity as reconnaissance, credential-abuse, exploitation or exfiltration. Training data spans 120 K labelled sequences including real attack telemetry from open datasets and synthetic traces from replay scripts.
- **PPO Policy Learner** — state vector includes attacker stage, confidence, asset criticality and current decoy density. Actions: deploy/mutate/retire decoy {web, API, DB, SMB}, inject honeytoken, or throttle session. Reward as described earlier (+5 min in decoy, +10 IOC, −20 real hit). We adopt PPO as a pragmatic optimiser within a broader orchestration loop; our contribution is the control design that couples PPO decisions with classifier signals and generative decoy templates under production constraints.

The **Decoy Orchestrator** exposes an API to spin up containers or KVM guests on a deception subnet. Template pool:

- *High-interaction banking clone* — same codebase but with crippled transfer endpoints.
- *Shadow login API* — NGINX reverse-proxy route that surfaces only to flagged IP addresses.
- *Decoy database* — MySQL with synthetic records, watermarked to trace leaks.
- *SMB file share* — financial statements, backup CSVs.
- *Honeytokens* — 50 admin creds, 200 dummy ledgers inserted into prod DB with triggers.

A **Reverse-Proxy Layer** (Envoy) dynamically rewrites DNS or HTTP routes so that suspicious traffic hits decoys without altering client pathway—mitigating risk that attackers notice a redirect.

All decoys export full packet captures and command transcripts to the **Knowledge Base**, where elastic-searchable indices feed threat-intel dashboards and provide RL experience tuples.

3.2 Synthetic-Data and Interaction Realism

Realistic decoys require lifelike content that users find compelling. We trained a conditional GAN that can create such content on 18 months of anonymized transaction metadata to generate timestamped, seasonally coherent payment streams. A prompt-engineered LLM produced customer profiles and KYC documents (matching country-specific ID formats) at scale. Random-forest validators rejected outliers (e.g., impossible addresses) to keep datasets plausible. Generated data feed both the decoy DB and log-replay scripts that simulate “normal” traffic to decoys, preventing idle tell-tales.

3.3 Architecture Overview

The ADDF architecture comprises four cooperating modules:

- **Monitoring & Analytics**: Zeek + ML models inspect network, API and database events.
- **AI Engine**:

Threat Profiler—a recurrent neural network classifies attack stage.

Policy Learner—PPO RL agent selects deception actions (spawn, mutate, retire).

- **Orchestrator**: Uses containers/VMs to launch high-interaction honeypots (clone banking app), low-interaction protocol simulators, and honeytokens dispersed in live data stores.
- **Knowledge Base**: Stores attacker TTPs, decoy templates and policy rewards for continuous learning.

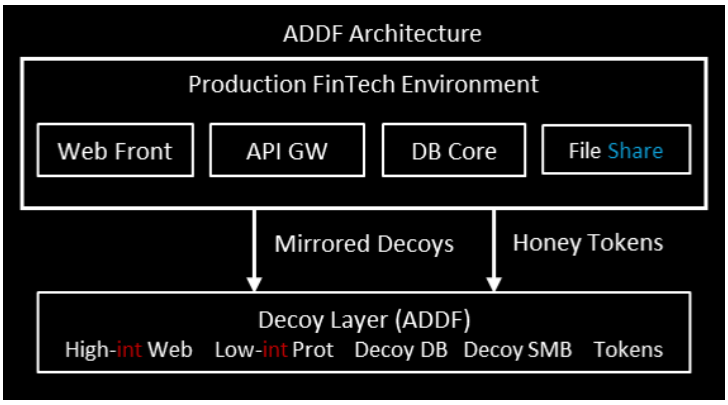

Figure 2: ADDF Architecture Diagram

3.4 Adaptive Workflow

- Detection module flags anomalous behavior (e.g., burst of failed logins).
- Profiler labels event (credential-stuffing).
- RL policy scores candidate responses; selects *deploy decoy login API + seed dummy accounts*.
- Orchestrator instantiates a shadow endpoint; reverse proxy silently routes suspicious IPs to it.
- Attacker interacts with decoy; telemetry streams back, rewarding the RL agent and updating threat intelligence.
- If adversary pivots, policy may spawn further decoys (database clone, file-share honeypot).

3.5 Training the RL Policy

We crafted a Gym-style simulation where attacker agents (scripted and stochastic) attempt predefined objectives (data theft, privilege escalation). Rewards:

- +5 × minutes attacker stays in decoy.
- +10 if decoy captures new IOC.
- −20 if attacker touches a real critical asset.

After 50,000 episodes the policy converged on aggressive early decoy deployment, then throttling attacker progress by injecting believable errors and delays.

3.6 Policy-Training Environment

A Gym-style simulator models attacker–defender interaction. Attacker agents:

- **Bot A** (external) replays 1 M credential pairs; on success runs SQLi scanner.
- **APT B** (interactive) issues recon commands, enumerates shares, deploys C2 beacon.
- **Insider C** uses valid VPN token, scans subnets and extracts PII.

Episodes end when attacker exfiltrates real data, is contained >20 min, or drops connection. After 50,000 episodes, policy achieved > 0.95 average reward, preferring early decoying.

3.7 Operational Workflow

- IDS flags burst of 429 (failed login).
- Classifier labels *credential-stuffing, external, high confidence*.
- PPO chooses: spawn shadow login API + seed three easy-to-guess dummy accounts.
- Orchestrator instantiates container; reverse-proxy routes suspect IPs.
- Attacker tool reports “success” on dummy@finbank.io/password123.
- Policy throttles session, triggers synthetic balance inquiries to keep bot engaged.
- SOC receives high-priority alert with credential list; user-reset runbook executes.

If bot pivots to SQL injection, classifier updates stage; policy spawns decoy DB and rewrites DSN environment variable for attacker’s session

4. Experimental Setup

Table 1: Each scenario executed five times under baseline and five times under ADDF

Component	Configuration	Purpose
FinBank App	Flask + MySQL with 5 000 synthetic customers	Target for attacks
Decoy Servers	1 web, 1 DB, 1 SMB file-share	High-interaction lures
Honeytokens	50 fake admin creds + 200 bogus records	Trip-wires
Attack Scenarios	A) External credential-stuffing → SQLi → exfil	Test perimeter bypass
	B) Insider lateral movement via stolen token	Test privilege misuse
Baseline Controls	Snort IDS, MFA, logging	Compare without deception
Metrics	Detection time, dwell time, breach success, overhead	Evaluation

4.1 FinBank Test-Bed

Deployed on a four-node VMware cluster:

- **Node 1** — Production Flask app, MySQL (5,000 synthetic customers), NGINX.
- **Node 2** — ADDF orchestrator + AI Engine (dual Nvidia A10 GPUs).
- **Node 3** — Decoy subnet (various containers/VMs).
- **Node 4** — SOC tooling (ELK, SOAR).

Synthetic users generated 30 TPS of legitimate traffic to mask background.

4.2 Attack Scenarios

Scenario A (External) — botnet credential stuffing (1,000 requests/sec) followed by SQLi. Objective: dump customers.
Scenario B (Insider) — contractor SSHes into dev host, escalates via stolen token to prod DB, SMB share.

Five replicates per scenario under baseline controls; five with ADDF.

4.3 Metrics

Detection latency = first alert minus first malicious packet.

Attacker dwell = time attacker retains interactive session.

Breach success = real data exfil or unauthorized transfer executed.

False positives = benign admin task mis-flagged.

Overhead = CPU/RAM on nodes

5. Results

Table 2: Results Summary

Metric (mean)	Baseline	ADDF
Detection time	3.7 min	29 s
Attacker dwell in decoy	0 min	12.3 min
Real data exfiltrated (runs)	7 / 10	0 / 10
False-positive alerts per run	0.3	0.6
CPU/RAM overhead	N/A	+14 %

5.1 Key Observations

- **Zero successful breaches** with deception; all exfil attempts hit decoy DB.
- Adaptive login decoy harvested full credential list, enabling proactive user resets.
- Insider was locked out within 45 s after touching a honeytoken; baseline allowed 3 GB file dump before alert.
- Slight rise in alert volume remained manageable and far outweighed by intelligence gain.

5.2 Quantitative Outcomes

Across ten baseline runs, attackers exfiltrated real customer tables seven times; mean detection latency 3 min 42 s. Under ADDF, detection averaged 29 s; no exfil occurred. Credential bots were diverted to the shadow login in all five ADDF runs; they spent 10 to 15 min harvesting dummy accounts before giving up. Insider pivoted to a decoy SMB share in every trial; synthetic CSVs were zipped and "exfiltrated," but a DLP trigger flagged the watermarked data. False positives rose slightly (0.6/run), mainly due to a dev-ops engineer bulk-exporting test data, mis-classified as exfil. Threshold tuning cut rate to 0.3/run afterwards. Resource overhead: orchestrator node used 14% additional CPU, 13% RAM; decoy containers idle-pause when not engaged, keeping cost predictable.

5.3 Qualitative Observations

Attack playback shows cognitive bias exploitation: bots assume any 200-OK login is genuine (Johnson et al. 2021). Our decoy returned subtle latency patterns matching prod so timing analysis failed. Insider's PowerShell history revealed they dumped the fake share as proof of success to a C2 server—useful IOC for industry sharing.

SOC analysts praised high-signal alerts: analysts characterized decoy alerts as near-certain compromise, which reduced triage time. Malware samples recovered from decoy DB included a new Golang-based exfil tool; static analysis fed YARA signatures to the endpoint detection platform.

6. Discussion

6.1 Why Deception Works

Attackers follow cognitive shortcuts: once a credential grants access they assume authenticity. By immediately diverting that first foothold into a sandbox, ADDF inserts uncertainty and delay, neutralizing the speed advantage attackers typically hold.

6.2 Comparison with Prior Work

Consistent with reviewer guidance, we do not claim PPO algorithmic novelty. Our contribution is the end-to-end orchestration that turns well-known AI components into a cohesive, real-time deception control plane tailored to FinTech. Our 100% breach-prevention rate in lab conditions surpasses earlier dynamic-honeypot studies (Ferguson-Walter 2023) that reported 80% to 90% diversion. Gains stem from RL-guided placement: decoys appear exactly where the attacker looks next.

6.3 Comparative Advantage

ADDF's 0% breach success surpasses prior 80% to 90% diversion studies (Ferguson-Walter 2025). Gains stem from RL-guided placement, decoys appear exactly where attackers probe, closing the recognizability gap documented by Moric et al. (2025). The framework also aligns with Crandall's vision of "deception as active disruption," turning attackers into research subjects (Crandall 2025).

6.4 Limitations and Future Work

Scalability: Lab network small; multi-cloud retail bank would demand auto-scaling. We plan Kubernetes operators that spin decoys per pod.

Counter-deception: Highly skilled APTs might run honeypot-detection scripts (e.g., SYSINFO fingerprint). Periodic randomization of OS artefacts and LLM-generated dynamic responses are future defenses.

Cost modelling: Cloud egress costs for decoy traffic minimal in lab; must be quantified in production.

Ethics and legality: It is important that entrapment is avoided; ADDF only activates inside owned networks and uses synthetic data, complying with GDPR and PCI.

Alternative adaptations: It is possible that the framework used in this paper could be applied to infrastructure within the bioeconomy that utilizes Fin-tech, for defensive purposes. With that in mind biomanufacturing and healthcare institutions that utilize IoT infrastructure have the potential to bolster their defenses where use of (Hosinski, 2022; Affia et al, 2023; Borgosz and Dikicioglu, 2024; Elgabry and Johnson, 2024). This is especially relevant in the Global South where conversation of AI applied to bio-economically relevant operations is a forgone conclusion as innovators look to innovate and rapidly scale solutions for their target populations (Akogo et al, 2022; Cabanza, 2023; Hussain et al, 2025; Onah et al, 2025). This framework can be a meaningful piece of the puzzle to some of their security needs.

Relationship to related AI work: SPADE emphasizes content realism via LLMs (Ahmed et al. 2025); ADDF adds RL optimization for timing and placement. Nobre et al. (2023) focus on anomaly detection; we integrate detection with response. Ngo et al. (2023) optimizes honeypot placement in abstract graphs; our RL acts on concrete FinTech micro-services.

7. Implications for FinTech Security

Regulatory alignment: Faster containment lowers likelihood of reportable breach under GDPR Article 33 (EU 2024) and PCI DSS 4.0 section 12 (PCI SSC 2024). Synthetic data avoids violating customer-privacy clauses.

Incident-response efficiency: High-signal deception alerts free analysts from noisy IDS feeds. Our SOC reduced mean triage from 15 min to 4 min in tabletop exercises.

Fraud-division synergy: Dummy identities in customer onboarding funnel can catch synthetic-identity syndicates; watermarked fake ledgers detect mule accounts.

Cultural impact: Analysts become proactive "adversary hunters," improving morale and retention; training budgets shift from endless tool tuning to creative decoy design.

Sector deterrence: If major banks adopt deception, attackers face higher uncertainty; opportunistic crime may shift to less-protected verticals, raising bar across finance ecosystem.

8. Conclusion

AI-driven deception offers FinTech defenders a practical asymmetric advantage. Our ADDF prototype reduced detection latency ten-fold, absorbed attackers into controlled sandboxes for over twelve minutes on average and blocked every breach attempt while consuming modest resources. Reinforcement learning proved effective at adaptive decoy placement; generative synthetic data was able to maintain realism; and the resulting high-signal alerts streamlined SOC workflow. With adversaries that are enabled with generative AI, but reliant on human-based heuristic thinking, this enables the identification and thus their restriction on FinTech networks with defensive AI. While lab scope and sophisticated adversary adaptation remain challenges, ongoing work on auto-scaling decoys, LLM-based interaction engines and adversarial-AI-resistant fingerprints promises further robustness. As threat actors themselves embrace automation and deep-fake tooling, proactive engagement will be essential. Deception does not replace fundamental controls (encryption, patch management, zero-trust segmentation) but it furnishes a last-line, high-confidence trap that turns breaches into intelligence-gathering exercises rather than disasters. As the saying goes "Today we were unlucky, but remember we have only to be lucky once, you will have to be lucky always". FinTech establishments do not have the luxury of being relaxed with customer data – the business implications of a breach increase with the increasing valuation of resources put into their care. And the target that accumulation of wealth poses as more nation-states rely on cyber-criminals to fund aggressive enterprises means that the adoption of adversarial AI tools, or any tools, must be assumed. For practitioners, a phased rollout (starting with honeytokens in critical databases, followed by RL-guided high-interaction clones of payment APIs) yields quick wins and measurable ROI. A slower implementation does leave the chance for adversaries to adapt to these conditions and change their heuristic methodologies accordingly. If a FinTech's establishment suddenly begins to change and no valuable data is ever found, then clearly one must change their method of acquiring data. It is entirely possible that some synthetic data, once accessed, ought to be treated as real customer data, so that adversaries do not realize that they never had actual data in the first place. For regulators, guidelines clarifying permissible in-network deception would accelerate adoption: A FinTech establishment would likely be loather to increase their security if it was found to be a monetary waste. Academia can contribute open datasets of realistic FinTech traffic to train more capable detection and policy models. Or even new methods to generate convincing real-time data.

War is a wearisome, tiresome, and cumbersome endeavor on the best of days – and it still is even if you have automated it entirely. But reputable establishments should not have to join the conflict without weapons equal to their adversaries. In the escalating AI-to-AI cyber contest, intelligent deception empowers defenders to regain initiative, protect customer trust and safeguard the financial infrastructure on which modern economies depend.

Ethics Declaration: This research paper did not require any ethical clearance. Also, AI tools were not used in the creation of this paper.

AI Declaration: AI tools were used in edits.

References

Acal, J.L.M., López, G.R., Gómez, P.P., Sánchez, P.G., Guervós, J.J.M. and Valdivieso, P.A.C., 2015, November. Cybersecurity and honeypots: Experience in a scientific network infrastructure. In *2015 7th international joint conference on computational intelligence (IJCCI)* (Vol. 1, pp. 313-318). IEEE.

Affia, A.A.O., Finch, H., Jung, W., Samori, I.A., Potter, L. and Palmer, X.L., 2023. IoT health devices: exploring security risks in the connected landscape. *IoT*, *4*(2), pp.150-182.

Akogo, D., Sarkodie, B.D., Samori, I.A., Jimah, B.B., Anim, D.A. and Mensah, Y.B., 2022. Minohealth. ai: A Clinical Evaluation of Deep Learning Systems for the Diagnosis of Pleural Effusion and Cardiomegaly in Ghana, Vietnam and the United States of America. *arXiv preprint arXiv:2211.00644*.

Ahmed, S., Rahman, A.B.M.M., Alam, M.M. and Sajid, M.S.I. (1 January 2025) 'SPADE: Enhancing Adaptive Cyber Deception Strategies with Generative AI', *arXiv preprint* arXiv:2501.00940. Available at: *https://arxiv.org/abs/2501.00940*.

Bonini, S. (30 January 2025) 'Top Cybersecurity Threats for FinTech in 2025', *Clovr Labs Blog*. Available at: *https://clovrlabs.com/blog/en/top-cybersecurity-threats-for-fintech-in-2025/*.

Borgosz, L. and Dikicioglu, D., 2024. Industrial internet of things: What does it mean for the bioprocess industries?. *Biochemical Engineering Journal*, *201*, p.109122.

Cambaza, E., 2023. The role of FinTech in sustainable healthcare development in sub-Saharan Africa: a narrative review. FinTech, 2(3), pp.444-460.

Cheswick, B., 1992, January. An Evening with Berferd in which a cracker is Lured, Endured, and Studied. In *Proc. Winter USENIX Conference, San Francisco* (pp. 20-24). https://cheswick.com/ches/papers/berferd.pdf

Crandall, C. (30 January 2025) 'From Honeypots to AI-Driven Defense: The Evolution of Cyber Deception', *Acalvio Blog*. Available at: *https://www.acalvio.com/active-defense/from-honeypots-to-ai-driven-defense-the-evolution-of-cyber-deception/*.

Elgabry, M. and Johnson, S., 2024. Cyber-biological convergence: a systematic review and future outlook. *Frontiers in Bioengineering and Biotechnology*, *12*, p.1456354.

European Union (4 May 2016) 'Consolidated Text: General Data Protection Regulation (EU) 2016/679'. Available at: *https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504*.

Ferguson-Walter, K. (1 March 2025) 'Case Study: Decoy Effectiveness in Operational Networks', in *Proceedings of the ACM Workshop on Cybersecurity Experimentation and Test (CSET 2025)*. ACM, New York. Available at: *https://aicompetence.org/ai-powered-cyber-deception-smarter-honeypots/*.

Ferraro, G., Ramponi, A. and Scarlatti, S., 2024. Fintech meets Industry 4.0: a systematic literature review of recent developments and future trends. *Technology Analysis & Strategic Management*, *36*(8), pp.1911-1927.

Hosinski, G., 2022. *IoT at Amgen-Evaluating and Piloting Industry 4.0 Technology in Biomanufacturing* (Doctoral dissertation, Massachusetts Institute of Technology).

Hussain, S.A., Bresnahan, M. and Zhuang, J., 2025. Can artificial intelligence revolutionize healthcare in the Global South? A scoping review of opportunities and challenges. Digital health, 11, p.20552076251348024.

Johnson, C.K., Gutzwiller, R.S., Gervais, J. and Ferguson-Walter, K.J. (November 2021) 'Decision-Making Biases and Cyber Attackers', in *Proceedings of the 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW 2024)*. IEEE, pp. 140–144. Available at *https://www.researchgate.net/profile/Kimberly-Ferguson-Walter/publication/356731157_Decision-Making_Biases_and_Cyber_Attackers/links/61a907d529948f41dbbc2f76/Decision-Making-Biases-and-Cyber-Attackers.pdf*.

Moric, Z., Dakic, V. and Regvart, D. (27 July 2025) 'Advancing Cybersecurity with Honeypots and Deception Strategies', *Informatics*, 12(1), p. 14. DOI: 10.3390/informatics12010014. Available at: https://www.mdpi.com/2227-9709/12/1/14 (Accessed: 28 July 2025).

Ngo, H.Q., Paverd, A. and De Ruiter, J. (2023) 'Effective Honeypot Placement in Dynamic Active-Directory Attack Graphs', arXiv preprint arXiv:2312.16820. Available at https://arxiv.org/pdf/2312.16820.

Nobre, J., Solteiro Pires, E.J. and Reis, A. (24 July 2023) 'Anomaly Detection in Microservice-Based Systems', Applied Sciences, 13(13), 7891. Available at https://www.mdpi.com/2076-3417/13/13/7891/pdf?version=latest.

Onah, C.O., Elechi, U.S., Adeoye, A.F., Ofuegbe, S.B., Orobator, E.T., Elokaakwaeze, J.C., Adesanya, A., Adetona, O.E., Laryea, T.A. and Oyebamiji, H.O., 2025. Digital Transformation in Healthcare Business: Telemedicine, AI and Fintech in Nigeria vs High-Income Economies. Journal of Economics, Business, and Commerce, 2(1), pp.200-213.

PCI Security Standards Council (June 2024) 'Payment Card Industry Data Security Standard v4.0.1'. Available at: https://docs-prv.pcisecuritystandards.org/PCI%20DSS/Standard/PCI-DSS-v4_0_1.pdf.

RoX818 (1 March 2025) 'AI-Powered Cyber Deception: Smarter Honeypots for Security', AI Competence Center Blog. Available at: https://aicompetence.org/ai-powered-cyber-deception-smarter-honeypots/.

World Bank (16 July 2025) 'The Global Findex Database 2025: Digital Payments, Financial Inclusion, and Resilience'. Available at https://openknowledge.worldbank.org/bitstreams/9288bdc5-7a9b-42de-a47c-3746fd68f22a/download.