hackerone

eBook

Industry Spotlight:

Financial Services and AI Security

Industry Supplement to “Securing the Future of AI”

hackerone

Why it Matters

In a highly regulated, data-intensive sector like financial services, the introduction of generative AI adds a volatile new dimension to cyber risk. Financial institutions are already prime targets for cybercriminals—and with AI, attackers now have unprecedented tools to automate fraud, exploit customer trust, and bypass authentication.

HackerOne's agentic AI system, Hai, powered by Amazon Bedrock, arms financial services security teams with a team of agents to streamline triage, surface hidden risk patterns, and accelerate remediation—ensuring AI deployments remain secure, compliant, and production-ready inside AWS environments.

hackerone

Financial services

The Threat Landscape Keeps Getting Bigger

As financial institutions race to harness the power of AI, they're also inheriting a rapidly evolving threat landscape—one that's smarter, faster, and harder to detect. From real-time deepfakes to AI-crafted phishing campaigns, the risks are no longer theoretical, they're happening now. HackerOne combines structured AI red teaming, AI-specific pentesting, and the expertise of the world's largest community of ethical hackers, including many AWS-certified professionals, to address threats like these before they cause harm.

1 Deepfake-Enabled Wire Fraud & Synthetic Identity Attacks

Executives in financial institutions are at high risk of AI voice cloning or deepfake exploitation. In one real-world scenario, attackers used AI-generated audio to impersonate a CFO and authorize a wire transfer.

AI deepfakes pose material financial risk and introduce hard-to-detect insider threat vectors.

2 Prompt Injection in AI-Powered Training Assistants or Chatbots

With more banks rolling out AI-powered customer support or trade execution bots, the risk of prompt injection leading to unauthorized actions or data leakage becomes mission-critical.

3 AI-Enhanced Social Engineering and Phishing Attacks

Generative AI enables cybercriminals to craft convincing phishing emails,deepfake internal memos, or simulate live chats with bank personnel. These socially engineered attacks can bypass traditional red flags, especially when AI mimics internal jargon, writing style, or known workflows.

What used to take a hacker hours to craft now takes minutes with AI—making every inbox a high-risk attack surface.

4 Compliance Breakdown through Unvetted AI Models

Regulators are tightening scrutiny around AI's use in loan origination, credit decisioning, and fraud detection. The absence of adversarial testing or explainability could result in noncompliance with SEC, FINRA, or OCC guidelines.

Real-world threat

Deepfake CFO Heist in Financial Services

In early 2024, a finance employee at a multinational firm in Hong Kong was tricked into transferring

$25 million

during a video call. The twist? Every person on the call—including the CFO—was a deepfake.

The attackers used AI-generated voices and video to mimic leadership.The employee, believing it was a legitimate internal request, executed 15 wire transfers to multiple accounts.

The fraud wasn't discovered until days later when the employee contacted headquarters.

This case shows how AI voice and video deepfakes now bypass traditional phishing red flags and social engineering defenses—especially in industries like finance, where high-trust, high-value transactions are routine.

Source: https://blog.wellsins.com/corporate-case-study-25-million-deepfake-scam-sends-a-wake-up-call-to-corporate-cybersecurity

hackerone

AI Amplifies Third-Party & Supply Chain Risk in FinServ

Financial institutions rely heavily on a vast network of third-party vendors—cloud providers, fintech APIs, CRM platforms, data processors, and even LLM-based tools—to power their digital operations. As AI becomes embedded in these systems, the attack surface expands beyond direct control. HackerOne can run production-safe security tests directly in AWS environments—no data replication required—providing real-world validation across third-party integrations.

Unvetted AI inside third-party apps

Vendors may integrate LLMs or GenAI features into their tools without disclosing the risks or safeguards.

Training data leakage

Third-party platforms using customer logs or financial data to train their models may inadvertently expose PII or confidential trading logic.

Model poisoning or backdoor injection

If a vendor's AI supply chain is compromised, downstream institutions may unknowingly integrate tampered or biased models into critical systems.

Shadow AI risk

Business units may adopt AI tools independently of security governance, increasing blind spots.

As AI tools proliferate across vendors, contracts alone aren't enough. FinServ firms must extend adversarial testing, audits, and AI governance to their supply chain.

hackerone

Governance & Compliance Priorities

As financial institutions adopt AI, they must align model development and deployment with a complex web of regulatory, operational, and ethical expectations. The table below outlines key governance priorities and how adversarial testing and red teaming can support compliance at scale.

Priority Area	AI Security Risk	Governance Solution
Model Drift Monitoring	Unauthorized changes to behavior of deployed models.	Structured AI red teaming & continuous behavioral testing detect drift before it impacts production.
Audit Readiness	Lack of explainability & auditability for model decisions.	Integrate AI testing with internal audit workflows.
Third-party Vendor Risk	AI code or models embedded in FinTech APIs, SaaS platforms, or 3rd party LLM tools without proper vetting.	Extend AI-specific pentesting to third-party models & APIs, leveraging AWS-native integration and production-safe testing in live AWS environments for real-world validation.
Privacy Compliance (GLBA, GDPR, CCPA)	Training data may contain sensitive or personally identifiable information.	Apply masking, encryption, and audit trails on model training pipelines; and AWS-native workflows for continuous compliance.

Board Level Risk and Reputational Exposure

For financial institutions, AI missteps aren't just security issues, they're enterprise risks. A single AI-driven breach, hallucinated financial output, or compliance failure can result in regulatory penalties, investor backlash, and loss of customer trust. Boards are increasingly asking pointed questions about AI governance, model accountability, and the safeguards in place to prevent reputational harm. Risk, compliance, and security leaders must be ready with clear answers, and a roadmap for responsible AI deployment.

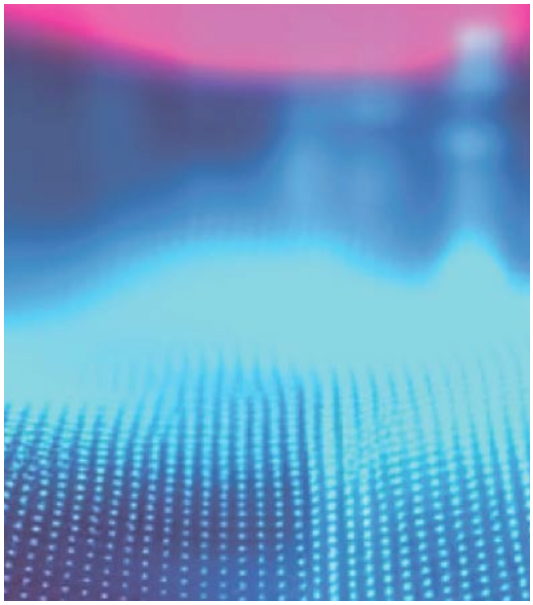
hackerone

Actionable checklist

Is Your FinServ AI Secure?

As AI accelerates innovation in financial services, the pressure to deploy responsibly—and defensively—has never been greater. Securing AI isn't just a technical challenge; it's a business-critical mandate tied to trust, transparency, and long-term resilience.

To help security, risk, and compliance teams assess their readiness, here's a focused checklist of key controls and questions every financial institution should apply to their AI systems—internally and across third-party vendors.



- **Have your GenAI models undergone adversarial testing?**
- **Are customer-facing LLMs hardened against prompt injection and hallucination?**
- **Can your AI system decisions be audited and explained to regulators?**
- **Is third-party AI usage inside your FinTech integrations monitored and red teamed?**
- **Is AI training data masked, compliant, and free of personally identifiable information?**

HackerOne helps financial institutions secure AI with human-led red teaming, AI-specific pentesting, and Hai-powered vulnerability management—integrated with AWS to protect innovation without slowing it down.

Get Started with HackerOne on AWS