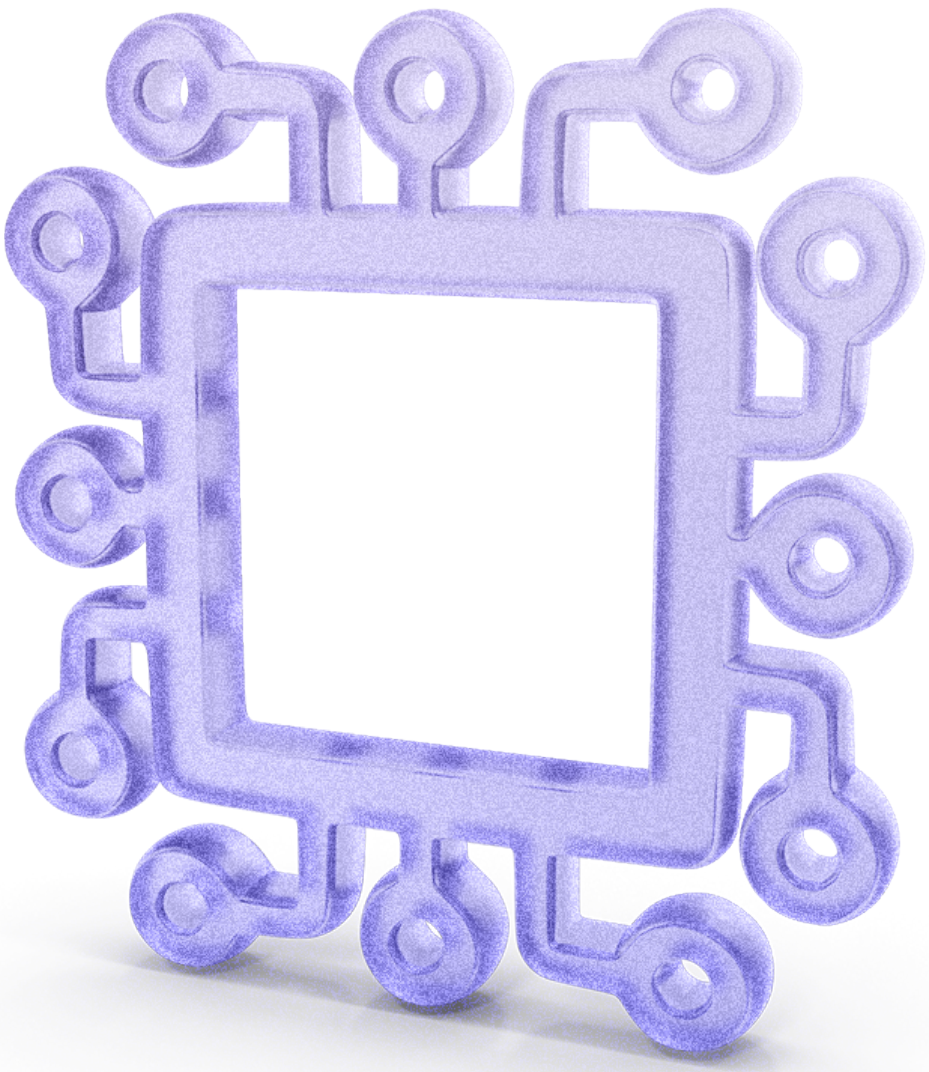
FS-ISAC

Financial Services and AI: Leveraging Opportunities, Managing Risks



February 2024

Contents

Executive Summary

Artificial intelligence (AI) promises business breakthroughs in the financial services industry. Banks use AI-based chatbots to answer customers' questions. Investment firms make predictions with AI-backed data. Insurance companies detect scams with AI-driven alerts. Boards of all types get insight from AI-written reports. From the back office to the front desk, the sector's AI applications are generating revenue, reducing costs, and expanding the scope of the business.

Yet the business impact promised by AI comes with significant risks to financial services firms. Chatbots can be altered to misinform customers. Predictive data can be poisoned. Fraud can be disguised. Board reports can be garbled, biased, or wrong.

Across the sector, adversaries are exploiting AI to rob, swindle, or corrupt financial services organizations.

The threats and risks associated with AI are difficult to detect and mitigate – and cyber criminals' methods constantly evolve. This has accelerated financial services' risk environment, challenging cyber teams to both defend their firms' security and resiliency while simultaneously empowering their organizations to capitalize on AI systems.

Governments and organizations around the world have responded to these cybersecurity issues with unusual urgency and depth. Over the last year or so, we saw the European Union AI Act[1], the White House Executive Order on AI[2], the Association for Computing Machinery Generative AI Guidelines[3], the Adversarial Machine Learning - Taxonomy and Terminology of Attacks and Mitigations from the National Institute of Standards and Technology (NIST)[4], US CISA and UK NCSC Joint Guidelines for Secure AI System Development[5], NIST's AI Risk Management Framework[6], OWASP's Top 10 for LLM Applications[7], ACSC's Engaging with Artificial Intelligence[8], and many other standards, regulations, guidelines, and advisory publications were released.

The FS-ISAC AI Risk Working Group formed to analyze the threats and opportunities inherent in AI and provide additive resources that build on the expertise of government agencies, standards bodies, academic researchers, financial services partners such as FSSCC[9] and BPI/BITS[10], and the knowledge of FS-ISAC members to mitigate risks, fortify the resiliency of financial services organizations, and safeguard the trust of customers, investors, and regulators.

The group's first body of work is a compendium of perspectives that help financial services institutions use AI securely, responsibly, and effectively. In six white papers, the FS-ISAC AI Risk Working Group provides practical frameworks and tactics that financial services firms can customize to their size, needs, and risk appetites according to each relevant function in the institution. The following table describes the key audiences within firms for each of the papers.

FS-ISAC AI Risk Working Group Publication	Business Function
Cybersecurity	
Adversarial AI Frameworks: Taxonomy, Threat Landscape, and Control Frameworks	Cybersecurity, Software Engineering, Information Technology
Building AI into Cyber Defense	Cybersecurity
Combating Threats and Reducing Risks Posed by AI	Cybersecurity, Fraud, Legal, Customer Service
Policy and Implementation	
Responsible AI Principles	Executive Leadership, Legal, Human Resources, Customer Service, Software Engineering, Data Privacy Office, Information Technology
The Generative AI Vendor Evaluation and Qualitative Risk Assessment	Procurement, Information Technology, Data Privacy Office
Framework of Acceptable Use Policy for External Generative AI	Executive Leadership, Legal, Human Resources, Data Privacy Office, Information Technology

These documents are point-in-time views that indicate the complexity and possibilities that AI introduces into our environments. The FS-ISAC AI Risk Working Group will continue to update and expand on these and other topics.

Cybersecurity

Adversarial AI Frameworks: Taxonomy, Threat Landscape, and Control Frameworks

This comprehensive paper enumerates, defines, and maps the existing threats associated with AI in financial services, as well as the unique and novel threats, weaknesses, and security controls related to generative AI usage and systems. Its Taxonomy of attacks aligns with and expands on the NIST Taxonomy of AI Risk Management and defines key terminology. The paper's technical explanations contextualize AI-related threats to provide a risk hierarchy.

By establishing common definitions of attacks, clarifying vulnerabilities, and relating effective controls, Adversarial AI Frameworks is a foundational work for the other five FS-ISAC AI Risk Working Group papers and for cybersecurity in the industry.

Building AI Into Cyber Defense

Focusing on opportunities for leveraging AI in cybersecurity and risk technology, this document highlights the key considerations, specific domains, and practical use cases that apply to financial services firms. In the highly regulated financial services sector, opportunities include:

- Anomaly detection
- Creating structure in unstructured data
- Empowering content generation
- Efficient data retrieval
- Open- and closed-source data tradeoffs
- Staff education
- Architectural design

By examining effective applications of AI-based solutions to cyber defense, this paper is a practical tool for financial services cybersecurity teams assessing the potential of AI solutions in their function.

Combating Threats and Reducing Risks Posed by AI

This paper describes the primary cyber threats and risks – both human and technological – that AI can precipitate in financial services institutions. The list of external and internal dangers include:

- Adversarial use of AI, such as malicious code, AI poisoning, deepfakes, and social engineering
- Adversarial targeting of AI-based solutions, including data poisoning, integrity attacks, and open-source compromises
- Risks related to AI models' data sourcing
- Inadvertent risks arising from the use of generative AI, such as hallucinations and unintentional bias
- External factors, including legal, regulatory, and ethical threats

Covering a broad spectrum of viewpoints concisely, the document provides detailed insight into AI's threat landscape and the best practices and mitigation approaches necessary to combat threats and reduce the risks associated with using AI.

Policy and Implementation

Responsible AI Principles

This paper helps financial services organizations consider principles and practices related to the responsible use of AI, with a close examination of data sources, governance structures, and transparency mechanisms in a holistic framework for the ethical deployment of AI. Though designed for the

financial services sector, the work is broadly applicable, covering:

- Safe, secure, and resilient AI systems
- Explainable and interpretable AI systems
- Privacy-enhanced AI systems
- Fairness with harmful bias managed in AI systems
- Ensuring valid and reliable AI systems
- Enhancing accountability and transparency in AI systems

The paper references and builds on the NIST AI Risk Management Framework, the White House Executive Order on AI, UNESCO Recommendations on Ethics of AI, and many other established risk management perspectives. Those approaches, as well as time-honored values and existing industry standards, are considered within the recommendations of ethical AI usage and development. By examining the principles intrinsic to the responsible use and management of AI, this paper empowers organizational alignment of AI practices with the industry's high standards of ethics and trustworthiness.

The Generative AI Vendor Evaluation and Qualitative Risk Assessment

The Vendor Generative AI Risk Assessment workbook is a customizable tool designed to help financial services organizations – of any size or type – assess, select, and make informed decisions about generative AI vendors while managing associated risks. The tool is designed to assist third-party risk programs with the discovery and evaluation of GenAI solutions and use cases within their organizations and supplement existing TPRM processes. The assessment can be tailored to accord with the organization's risk appetite.

The document expands upon industry best practices and frameworks, such as NIST's AI Risk Management Framework 1.0, NIST CSF, the FFIEC's IT Examination Handbook, and the FFIEC's 2023 Interagency Guidance on Third-Party Relationships: Risk Management.

Framework of Acceptable Use Policy for External Generative AI

Financial services firms can use this framework to design their own acceptable use policy for external generative AI and upgrade their security and risk management policies to incorporate safe and responsible AI use into their security programs.

Offering policy guidance on both permissive and stringent approaches, the tool helps organizations decide the right balance for themselves by providing guidance on considerations such as:

- Data provided to generative AI systems
  - Confidentiality
  - Responsibility
  - Access
  - Monitoring
- Data received from generative AI systems
  - Accuracy
  - Representation as the employee of company
  - Attribution

The document contains sample policy text incorporating acceptable use considerations (such as the CIA Triad) and other considerations specific to AI and external systems that firms can adapt for their own use.

Where to Focus First: Threat Priorities

Among the risks and threats discussed in the FS-ISAC AI Risk Working Group's analyses are a few concerns notable for their probability or prevalence. A summary of the most pressing – and sometimes most damaging – attack vectors are listed below.

Deepfake Attacks: Impersonating leaders, employees, and customers is an effective – and growing – method of attacking financial services companies. AI makes it more successful. Adversaries can readily obtain the tools to create convincing audio or video deepfakes to defraud financial services firms or degrade

their reputations. Institutions must combat these attack vectors cross-functionally, creating or reinforcing cybersecurity procedures across customer services, employee support help desks, cyber threat intel teams, and other cybersecurity functions.

Employee Use of GenAI: The risks associated with generative AI aren't all adversarial. Some arise from authorized use cases: developers may unknowingly use another firm's proprietary code, customer service agents may give incorrect information to customers, employees may inadvertently share PII for legitimate reasons. Yet prohibiting the use of GenAI can drive employees to use external tools – which may introduce greater threats – to succeed in their roles. Organizations need education and dialogue to determine safe uses of GenAI relative to the firm's risk posture.

New Phishing/BEC Techniques: GenAI enables adversaries to craft sophisticated phishing business emails, texts, and other communications – including deepfakes – without the tell-tale typos employees have long relied on to spot ill intent. Financial services organizations must therefore educate employees to assess malicious communications using established processes, updated perspectives, and behavioral – not grammatical – cues (e.g., 'Would the CEO actually contact me for that information or to perform that task?') in the absence of the traditional red flags.

Use of Proprietary, Copyrighted, or Erroneous Information: Financial services organizations may have little or no insight into the provenance of the data used by their large language models (LLMs) or large multi-modal models. As a result, their AI systems' outputs may include other organizations' IP, sensitive customer PII, or dangerously inaccurate or biased information. That puts institutions at risk of legal and reputational damage. Financial services organizations are increasingly responsible for understanding the provenance of LLM training efforts and vendors' data updates.

Adversaries Aren't the Only Threat: Bias, Hallucinations, and Ethical Usage

Financial services firms are also vulnerable to bias, hallucinations, and unethical usage.

Perhaps the best-known and least understood of non-adversarial threats are hallucinations, the presentation of non-factual information as though it were true.

Concerning bias, datasets that include insufficient or prejudiced information – as open-source datasets may do – can train LLMs to return inaccurate or inappropriate outputs. Among the most concerning of those outputs is hateful content, which can cause substantial reputational harm to the institution.

The FS-ISAC AI Risk Working Group details many mitigation techniques – chiefly in the *Adversarial AI Frameworks: Taxonomy, Threat Landscape, and Control Frameworks* and *Combating Threats and Reducing Risks Posed by AI* white papers – to combat bias. Two highly effective techniques are guardrails that identify harmful responses in GenAI, and human-in-the-loop validation that makes AI outputs recommendations, rather than directly performing actions, in classical and generative AI systems.

But a more amorphous problem persists: how can financial services organizations use AI safely, responsibly, and ethically? There is no singular answer to this problem; as discussed in the FS-ISAC AI Risk Working Group's white papers, financial services firms must judge the correct answers as they align to their culture, customers, and risk posture. However, there are business consequences related to the answer. For example, while responsible use helps avoid reputational harm, it also incorporates use-case planning and thorough documentation, which increases the efficacy of decisions made with AI outputs and provides mechanisms that increase resiliency during a cyber incident. Adherence to the principles of responsible AI use helps prevent financial losses due to improperly generated or inadequately applied AI outputs. Established, documented, and demonstrably principled AI practices

may help serve compliance needs as well. The FS-ISAC AI Risk Working Group's compendium of documents is a foundation that financial services organizations can use as they determine their internal principles and practices.

Conclusion

The FS-ISAC AI Risk Working Group's frameworks for mitigating AI risk are a response to an inflection point in the sector. So are the new legislation, policy documents, and advisory publications. For many years, the use cases for AI were narrow and specific. Now AI solutions are ubiquitous, broadly applicable, and incredibly powerful. They offer financial services organizations extraordinary efficiency and scope – as well as a wider attack surface and greater exposure to risk.

Perhaps the most dangerous risk is the loss of trust. Misapplied or insecure AI systems can easily erode the trust of customers, regulators, and investors. Their trust is fundamental to the financial services sector, and indeed the functioning of the entire global financial system.

Viewing AI within cybersecurity frameworks enables cyber teams to better assess the AI threat landscape and isolate gaps in their existing approaches. Using AI systems ethically, responsibly, and safely defends organizations across a spectrum of external, internal, and compliance issues in the AI risk ecosystem. And it sustains stakeholder trust in financial services institutions.

Ultimately, the FS-ISAC AI Risk Working Group's goal is to help the sector safely realize the potential of AI systems, avoid the risks – and help firms sustain the trust that is at the core of the industry.

References and Resources

1 https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

2 The White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

3 The Association for Computing Machinery's Principles for the Development, Deployment, and Use of Generative AI Technologies

4 National Institute of Standards and Technology (NIST) Adversarial Machine Learning - Taxonomy and Terminology of Attacks and Mitigations

5 CISA and NCSC Joint Guidelines for Secure AI System Development

6 National Institute of Standards and Technology (NIST) AI Risk Management Framework

7 Open Worldwide Application Security Project (OWASP) Top 10 for LLM Applications

8 https://www.cyber.gov.au/resources-business-and-government/governance-and-user-education/governance/engaging-with-artificial-intelligence

9 Financial Services Sector Coordinating Council

10 Bank Policy Institute

Contact

fsisac.com

media@fsisac.com