DARK ARTS

The Unguarded Frontier

Why Firms Are Losing Control of Customer Data in the Age of AI

A Dark Arts White Paper | April 2026

Executive Summary

Artificial intelligence is transforming financial services at unprecedented speed. Over 85% of financial institutions are now actively deploying AI across fraud detection, risk modelling, trading, and customer service. Global annual spending on AI in the sector exceeds $20 billion. The race to adopt is well underway.

But in the rush to capture competitive advantage, a critical vulnerability has emerged: the systems, policies, and checkpoints that protect customer data were not designed for this new paradigm. Legacy compliance frameworks built for structured databases and controlled access points are being overwhelmed by a technology that operates in fundamentally different ways.

This paper argues that financial services firms are moving too fast with AI adoption and are failing to establish sufficient guardrails to ensure customer data remains protected. The consequences are not hypothetical. **They are already materialising.**

---

1. The Speed Problem: Adoption Without Architecture

The financial services industry has adopted AI at a pace that far outstrips its governance infrastructure. A 2025 survey by Caspian One found that 70% of financial institutions were utilising AI at scale, up from just 30% in 2023. McKinsey reports that 91% of banking institutions have established centralised generative AI functions. Hedge funds, wealth managers, and insurers are all in the same race with 68% of hedge funds now employing AI for market analysis and trading strategies.

Yet the implementation reality tells a different story. Only 38% of AI projects in finance meet or exceed their ROI expectations, according to Deloitte. Over 60% of firms report significant implementation delays. The gap between ambition and execution is enormous and within that gap lies a growing data protection crisis.

The question is not whether financial services firms should adopt AI. It is whether they can do so without losing control of the data that their customers have entrusted to them.

The problem is structural. Most financial institutions have layered AI capabilities onto existing technology stacks without fundamentally rethinking their data governance. Customer records, transaction histories, account details, and personally identifiable information are flowing through systems and workflows that were never designed with large language models in mind.

---

2. The Shadow AI Crisis

Perhaps the most immediate and underappreciated risk is the phenomenon known as "shadow AI", employees using unsanctioned artificial intelligence tools in the workplace without organisational oversight or approval.

The numbers are striking:

- **78% of AI users at work** bring their own AI tools, a trend Microsoft has termed "BYOAI."
- **73.8% of ChatGPT accounts** used in workplace settings are non-corporate, meaning they lack the security controls of enterprise versions.
- **485% increase** in the volume of corporate data sent to AI tools between March 2023 and March 2024.
- **27.4% of data** uploaded by employees to AI tools was classified as sensitive, up from 10.7% just one year earlier.

In practical terms, this means that employees across the financial services sector are copying customer data, account numbers, transaction records, loan applications, investment portfolios, into consumer-grade AI tools to help them draft emails, analyse spreadsheets, summarise documents, and generate reports.

The Samsung Warning

While not a financial services firm, the Samsung incident of March 2023 remains the canonical illustration of this risk. In three separate instances, Samsung engineers uploaded proprietary source code and confidential meeting transcripts into ChatGPT's free web interface, data that was then incorporated into OpenAI's training pipeline.

For a bank or insurer, the equivalent scenario is an analyst pasting customer portfolio data into ChatGPT to generate a summary, or a compliance officer feeding sensitive regulatory correspondence into Claude to draft a response. The data leaves the controlled environment, enters a third-party system, and if the employee is using a free or personal account, may be used to train the model itself.

The Training Data Question

This is where the risk becomes acute. Consumer-tier AI services operate under fundamentally different data handling policies than their enterprise counterparts:

- **OpenAI ChatGPT (free tier):** Conversations may be used to train and improve models.
- **Anthropic Claude (free/Pro tiers):** As of late 2025, data from free, Pro, and Max accounts may be used for training unless the user explicitly opts out.
- **Enterprise versions:** Both OpenAI and Anthropic prohibit training on customer data by default for enterprise and API clients.

The distinction is critical. When an employee uses their personal AI account at work, they are effectively opting their employer's customer data into a training pipeline. Once incorporated into model weights, there is no mechanism for recall. The data cannot be deleted, corrected, or retrieved. It becomes part of the model's learned knowledge, permanently.

3. People Are Not Being Trained

The shadow AI problem is not primarily a technology failure. It is a training failure. Employees are not being equipped with the knowledge they need to understand the data implications of the tools they are using.

The Verizon 2025 Data Breach Investigation Report found that 15% of employees routinely use generative AI on corporate devices. Separately, 33% of workers admit to sharing data on unapproved AI platforms, with 37% disclosing employee data and 24% sharing sales or financial data through these channels.

In most cases, these employees are not acting maliciously. They are trying to be more productive. They have discovered that AI tools can help them work faster, write better, and analyse data more efficiently. But they have not been told or have not understood that using a personal ChatGPT account to process customer data is fundamentally different from using an enterprise-grade tool with appropriate data protections.

The absence of clear AI usage policies, combined with the lack of practical training, has created a situation in which well-intentioned employees are the primary vector for data exposure.

This training gap extends beyond individual contributors. Many compliance teams and risk officers are themselves still developing fluency with AI technologies. They understand traditional data handling protocols, but they may not fully grasp how large language models process, retain, and potentially surface the information fed into them.

---

4. The Regulatory Reckoning

Regulators are paying attention. Across multiple jurisdictions and agencies, the regulatory framework around AI and data protection in financial services is tightening rapidly.

SEC: Examination Priorities and Enforcement

The U.S. Securities and Exchange Commission has made AI a focal point of its 2025 examination priorities. Examiners will assess registrants' policies and procedures for monitoring AI controls, fraud prevention, anti-money laundering compliance, and protection against client information loss.

The SEC has also moved from guidance to enforcement. In March 2024, it brought its first "AI-washing" actions against investment advisers Delphia and Global Predictions for making false claims about their use of AI. Combined penalties totalled $400,000. The SEC's Cybersecurity and Emerging Technologies Unit has stated that rooting out AI-related fraud is an "immediate priority."

NYDFS: Cybersecurity Guidance on AI

The New York Department of Financial Services issued targeted guidance in October 2024 identifying four primary AI cybersecurity risks: AI-enabled social engineering (including deepfakes), AI-enhanced cyberattacks, the expansion of data exposure through large AI datasets, and third-party AI tool vulnerabilities. Beginning November 2025, firms will be required to maintain complete and documented asset inventories of information systems and implement data minimisation practices, including the disposal of non-public information no longer needed for business operations.

GDPR and European Data Protection

The European Data Protection Board issued its opinion on AI models and GDPR compliance in December 2024, establishing that organisations must conduct documented, case-by-case legitimate interest assessments before using personal data for AI training. The standard is demanding: it must be "very unlikely" that individuals whose data trained the model could be directly or indirectly identified. For financial services firms operating across jurisdictions, this creates significant compliance complexity.

CCPA: New AI-Specific Provisions

California's Consumer Privacy Act has been amended with AI-specific provisions taking effect in 2025 and 2026. AB 1008 explicitly recognises that personal information can exist within AI systems. New regulations effective July 2025 require risk assessments for processing that presents significant privacy risk, with penalties of $2,500 per unintentional violation and $7,500 per intentional violation per affected consumer.

For a financial institution with millions of customer records, the arithmetic is sobering.

---

5. The Cost of Inaction

The financial and operational risks of inadequate AI data governance are not theoretical. They are quantifiable and growing.

- **Breach cost premium:** Organisations with high levels of shadow AI experience an additional $670,000 per breach, a 16% increase over firms with low or no shadow AI exposure.
- **Internal data leaks:** 46% of organisations have reported internal data leaks through generative AI applications.
- **AI-enabled fraud:** Deloitte projects that generative AI could propel U.S. fraud losses to $40 billion by 2027, up from approximately $12 billion in 2023.
- **Securities litigation:** AI-related securities class actions doubled between 2023 and 2024.
- **Deepfake attacks:** A 1,530% surge in deepfake attacks across Asia-Pacific between 2023 and 2024, targeting bank authentication systems.

Beyond direct financial exposure, firms face reputational damage that can erode client trust for years. In wealth management and private banking where relationships are the primary asset, a data breach involving AI mishandling of customer information can be existential.

---

6. What Prudent Firms Are Doing

The firms that are navigating this transition most effectively share several common characteristics. They are not avoiding AI; they are building the infrastructure to use it responsibly.

Establishing Clear AI Usage Policies

The starting point is a written, enforceable policy that defines which AI tools are sanctioned, how they may be used, and what data may and may not be processed through them. This policy must distinguish between enterprise-grade tools (with appropriate data handling agreements) and consumer-tier products (which should be prohibited for any work involving customer or proprietary data).

Investing in AI Literacy

Effective governance requires that employees at all levels understand the mechanics of AI data handling. This is not about banning tools, it is about ensuring that everyone from front-line advisers to C-suite executives understand the difference between an enterprise API call and a personal ChatGPT conversation, and why that distinction matters for their clients.

Deploying Enterprise-Grade Infrastructure

Where AI is deployed, it must operate within controlled environments with appropriate data isolation, access controls, audit trails, and retention policies. This often means migrating away from legacy systems that cannot support the segmentation and monitoring requirements of modern AI governance.

Building Continuous Monitoring and Audit Capabilities

AI governance is not a one-time implementation. It requires ongoing monitoring of data flows, regular audits of AI tool usage, and adaptive policies that evolve as the technology and regulatory landscape develop.

---

7. Closing the Gap

The financial services industry stands at an inflection point. AI offers genuine transformative potential, but only if the infrastructure supporting it is as sophisticated as the technology itself. The current trajectory, in which adoption dramatically outpaces governance, is unsustainable.

The firms that will emerge strongest are those that treat AI data governance not as a compliance burden, but as a competitive advantage, a signal to clients, regulators, and the market that they take their stewardship responsibilities seriously.

For legacy institutions, this often requires more than policy changes. It requires a fundamental re-examination of the technology stack itself, ensuring that systems can support data isolation, access controls, and audit capabilities that modern AI governance demands.

This is not a problem that can be solved with a memo. It requires architecture.

About Dark Arts

Dark Arts is a boutique software development firm that specialises in financial systems rescue and modernisation. We work with lenders, insurers, debt litigators, credit bureaus, fund managers, and other financial services organisations to modernise their technology infrastructure, ensuring it can support the compliance, security, and data governance requirements of an AI-enabled world.

If you are concerned about your organisation's AI data governance posture, or if you recognise the gap between your current systems and where they need to be, we should talk.

Contact us at info@darkarts.it **to schedule a confidential discussion about your AI readiness.**

Sources

1. CyberHaven, Shadow AI Report 2024
2. Microsoft Work Trend Index 2024
3. CIO Dive, "Samsung Employees Leaked Corporate Data via ChatGPT," April 2023
4. IBM Cost of a Data Breach Report 2024
5. NYDFS Industry Letter on AI Cybersecurity Risks, October 16, 2024
6. SEC Press Release 2024-36, March 18, 2024
7. Deloitte Center for Financial Services, 2024
8. CCPA AB 1008, effective January 1, 2025
9. SEC 2025 Examination Priorities
10. EDPB Opinion on AI Models and GDPR, December 18, 2024
11. Verizon 2025 Data Breach Investigation Report