Ivey | Publishing

W44898

GOVERNING INNOVATION: GOOGLE'S SOX CONTROLS FOR AI/ML IN FINANCIAL SYSTEMS

Glorin Sebastian and Eshan Bhatt wrote this case solely to provide material for class discussion. The authors do not intend to illustrate either effective or ineffective handling of a managerial situation. The authors may have disguised certain names and other identifying information to protect confidentiality.

Rapidly adopted in accounting and finance (through tools such as predictive analytics, document parsing, and chatbot-based automation), artificial intelligence (AI) and machine learning (ML) technologies had been transforming how organizations forecasted financials, managed risk, and executed operational tasks. While these tools improved efficiency, they also introduced risks especially when it came to financial reporting. These included data completeness and accuracy issues, model drift due to continuous learning, lack of transparency in decision-making, and challenges in reproducing results for audit purposes. This case examines how Google responded to these challenges by identifying gaps in its existing Sarbanes-Oxley (SOX) risk and control framework. In response, the internal risk and controls team developed and implemented a structured approach to make AI/ML models more auditable and better aligned with internal control standards. The results went beyond compliance. The framework helped reinforce broader principles of corporate governance—such as accountability, transparency, and oversight—within a rapidly evolving technical environment. What began as an initiative to meet the requirements of the Sarbanes-Oxley Act of 2002 (SOX) became a model for how controls could be extended to govern emerging technologies in financial systems. This case offers insight for leaders, compliance professionals, and boards seeking to integrate innovation with strong governance. It shows that with the right structure and proper controls in place, it is possible to keep pace with AI while still meeting the demands of governance, assurance, and regulatory trust.

Introduction

In early 2023, the finance leadership team at Google LLC (Google) faced a pivotal decision: how to scale its use of AI/ML models in financial forecasting without compromising regulatory compliance and internal control integrity. The goal was to enhance the accuracy of quarterly earnings forecasts, reduce manual workload, and generate data-driven insights at speed. However, traditional SOX controls—designed for rule-based systems and human-led processes—were not built to accommodate the complexities of models based on AI/ML technologies.

The introduction of adaptive AI/ML models using financial data brought challenges related to the model's transparency, auditability, and accountability—attributes that were essential for strong corporate governance. This case examines how Google's finance team responded, navigating the tension between

innovation and control, and ultimately building a scalable framework to support AI/ML integration while meeting SOX and broader governance requirements.

Problem Statement

The existing SOX control environment was generally optimized for static systems and deterministic logic. Enacted in 2002, SOX was a US federal law designed to improve the accuracy and reliability of corporate financial reporting by enforcing strict internal controls and audit requirements on publicly traded companies.[1] In contrast, AI/ML models deployed in finance—especially those used in duplicate invoice detection—introduced dynamic behaviour, learning from new data and evolving over time. During an initial risk assessment, the finance and compliance teams identified several concerns, including model reliability, fairness, interpretability, privacy, and security. However, when evaluated through the lens of SOX and financial reporting integrity, these risks required further refinement. The team aligned the identified risks with best practices outlined in the National Institute of Standards and Technology (NIST) AI Risk Management Framework to ensure regulatory coherence and control effectiveness.[2]

To understand the evolving risk landscape, it is important to distinguish between concerns, risks, and controls, particularly in the context of financial compliance. Under the Sarbanes-Oxley Act, internal controls are designed to ensure the accuracy, completeness, and reliability of financial reporting.[1] During the initial assessment, the finance and compliance teams identified issues related to AI/ML model behaviour, including transparency, fairness, and data integrity. These were assessed through a financial reporting lens and formalized as risks—defined as factors that could compromise auditability, consistency, or the integrity of reported outcomes. Each risk was then mapped to control measures aimed at prevention, detection, or mitigation. This progression from concern to risk to control reflects a core principle of internal control design in regulated environments.

Traditional SOX control environments were generally built for static systems with predictable, rule-based deterministic logic. AI/ML models, especially those used in areas such as duplicate invoice detection, behaved differently, adapting over time based on new data inputs. This adaptability introduced risks not typically encountered in conventional systems. During the review process, the finance and compliance teams worked to refine and categorize these risks in alignment with best practices from the NIST AI Risk Management Framework. Through iterative analysis and stakeholder reviews, including both lines of defense 1 and 2, several priority risk areas were defined as most critical for effective governance and control.

This risk classification and control mapping aligned with the widely adopted Three Lines of Defense model. In this approach, operational teams and model developers were responsible for owning and managing risks (Line 1), risk and compliance functions provided oversight and support (Line 2), and an internal audit offered independent assurance on the effectiveness of controls (Line 3).[3]

Through multiple iterations, the following key risks emerged as the most relevant for a controls framework. The below risk areas were documented based on risk ranking.

---

[1] US Securities and Exchange Commission, "Sarbanes-Oxley Act of 2002," accessed July 19, 2025, https://www.sec.gov/about/laws/soa2002.pdf.

[2] *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (National Institute of Standards and Technology of the US Department of Commerce, 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[3] *The IIA's Three Lines Model: An Update of the Three Lines of Defense* (The Institute of Internal Auditors, 2024), https://www.theiia.org/en/content/position-papers/2020/the-iias-three-lines-model-an-update-of-the-three-lines-of-defense/.

1. *Model drift*. As the model continued to update itself with new data, its output could shift and cause model anomalies or malperformance that were not in line with the model objective and design. This raised concerns about the reproducibility and consistency of forecasted financial data over time. From a compliance standpoint, it became essential to implement controls for monitoring model behaviour and documenting changes. Therefore, controls were introduced to ensure the AI/ML models were periodically tested to identify output drifts that could cause an adverse impact on financial outcomes.
2. *Robust documentation*. Ensuring a formalized record of the model's design logic, assumptions, training data lineage, or performance benchmarks was one of the key focuses of the controls and engineering teams. In scenarios when explainability was not explicit, mechanisms to increase explainability or accountability were implemented, including human intervention in outputs and increased documentation. Without this, internal and external auditors would have limited ability to understand or validate the model's behaviour, creating a gap in auditability and governance.
3. *Security and privacy*. Inappropriate privileged access to the AI/ML model could result in inappropriate modifications to system programs, algorithms, configurations, and data. The model handled sensitive financial inputs and was integrated into broader reporting systems. There was a growing need to implement strict access controls, encryption of sensitive model data, and tracking audit logs as well as access review controls to detect unauthorized access, manipulation, or data leakage.
4. *Data provenance*. This was identified as a key risk area since stale, bad quality, or biased datasets could impact the model safety and result in inaccurate outputs. Unvalidated data could result in the ingestion of malicious data. Source, method of acquisition, provenance, and choice of input data that trained the model was documented and defined before the model was deployed. Data processing and analysis was performed to identify noise, outliers, or erroneous data that could impact the completeness and accuracy of results.
5. *Bias and data integrity*. The reliability of the model was heavily dependent on the quality and representativeness of the data used for training. Concerns emerged that biased or incomplete data could introduce systemic errors into forecasts, potentially skewing financial decisions and disclosures. This reinforced the need for robust data validation and representational fairness mechanisms through diverse data used for training the models.
6. *Evolving testing protocols*. Recognizing that AI/ML models differed significantly from traditional software, the team identified the need to enhance validation and testing approaches. Existing quality assurance methods were expanded to include statistical performance checks, drift testing, and robustness evaluation under varied data scenarios. This shift toward standardized, AI-specific testing protocols improved confidence in the model's predictive reliability and long-term performance.

Context

The integration of AI/ML technologies into Google's financial models was driven by a strategic vision to enhance efficiency, harness emerging technologies, and align with evolving expectations in corporate governance. Finance leadership recognized a convergence of internal opportunities and external trends that made AI adoption both timely and valuable. The governance and compliance functions sought to establish readiness concurrently with the deployment and rollout of AI/ML models.

Industry-Led Innovation

As advanced AI/ML tools gained traction across the financial operations of peer companies, Google saw an opportunity to strengthen its internal forecasting and capital planning processes. The initiative wasn't reactive; it was strategic, aiming to improve accuracy, efficiency, and decision-making through enhanced analytical capability.

Anticipating Regulatory Scrutiny

Even in the absence of firm AI regulations, growing expectations from financial regulators emphasized the importance of model explainability, transparency, and auditability. By addressing these areas early, Google positioned itself to meet future regulatory standards while reinforcing investor trust.

Operational Efficiency Gains

Legacy financial workflows were heavily manual and time-consuming. By adopting AI/ML technologies, Google could streamline routine forecasting tasks, identify anomalies faster, and enable finance teams to focus on high-value strategic analysis.

Control Enhancement and Risk Mitigation

The dynamic nature of AI/ML models introduced new risks—such as model drift and data integrity issues—that traditional SOX controls were not designed to address. Google used this opportunity to implement stronger internal controls that aligned with emerging governance frameworks like the NIST AI Risk Management Framework.

Long-Term Strategic Differentiation

Beyond immediate efficiency, AI/ML integration was seen as a long-term lever for transformation. By embedding responsible AI practices early, Google aimed to differentiate its financial governance capabilities and create a scalable, compliant foundation for future innovation.

What began as a forward-leaning innovation effort quickly matured into a company-wide initiative to strengthen financial governance. Google's finance, risk, and audit teams collaborated to develop robust, SOX-aligned AI/ML controls that supported transparency and compliance, while still enabling technological progress.

Decision-Making Process

To address the governance and compliance risks AI/ML models introduced, Google's finance organization established a cross-functional team that brought together professionals from finance, risk management, audit and engineering teams. The group was empowered to act quickly while maintaining alignment with Google's internal control standards, aiming to set a precedent for how the company would govern AI/ML in financial reporting moving forward. The group's task was to do the following:

1. Identify risks specific to the AI/ML forecasting model's design, training, and use.
2. Develop and implement controls tailored to those risks, ensuring auditability and traceability.
3. Integrate those controls into Google's existing SOX framework without disrupting business agility and existing processes.

Approach

Risk identification. The team began with a structured risk assessment. This included both bottom-up technical reviews and top-down financial control mapping. The team ensured that the proposed risks aligned with the best practices the NIST AI Risk Management Framework outlined. Key considerations included the following:

- The opacity and adaptive nature of the model.
- Data dependencies and input quality.
- Potential downstream effects on reported financial metrics.
- Alignment with SOX principles on completeness, accuracy, and authorization.

Based on the identified risks, the team designed a set of information technology (IT) controls tailored to the AI/ML model lifecycle. The model lifecycle was divided into five phases. (See Exhibit 1 for details.)

Control design and mapping. The AI/ML lifecycle was segmented into five operational stages. Controls were developed for each stage and categorized by control type (preventative/detective) and risk materiality (low/moderate/high for SOX). (See Exhibit 2 for details.) The team also documented whether each control was manual or automated and its frequency (transactional or periodic). Automated controls were prioritized where possible to enhance efficiency and reduce the risk of human error.

Integration of AI/ML Controls into SOX

To ensure that the newly designed controls could withstand audit review, Google followed a disciplined integration approach:

- *Alignment with financial reporting risk*. Each AI/ML control was linked to a specific financial reporting risk, establishing a clear control-to-risk mapping. This gave both internal and external auditors visibility into how AI-related risks were being mitigated.
- *Testability and repeatability*. Controls were built to support consistent testing, with clearly defined control owners, documentation, and validation criteria. This helped ensure that audits could be performed across reporting cycles with confidence.
- *Evidence and traceability*. For every control activity, evidence was generated and stored. This included logs from model deployment tools, approvals from finance leadership, model training outputs, validation scripts, and results. These artifacts provided the foundation for traceable, end-to-end audit support.

The resulting framework not only supported SOX compliance but also advanced the company's broader commitment to governance over AI systems, setting a scalable precedent for future applications across the enterprise.

Challenges

As the control design effort progressed, the team at Google experienced a pivotal realization: Traditional SOX controls could not simply be adapted and applied to AI/ML environments. The dynamic, data-driven nature of these technologies required a fundamentally new approach, one that considered model evolution, input variability, and the absence of deterministic outcomes. This insight led to the creation of the AI/ML model lifecycle control framework (see Exhibit-1), which would eventually become a core component of Google's updated control strategy.

When the finance team at Google first explored integrating AI/ML models into its workflows, the team encountered an unexpected challenge: No one seemed to speak the same language. The data science group and Engineering teams spoke in terms of algorithms and precision scores, while the finance, risk and internal audit groups focused on compliance, controls, and risk. It quickly became clear that there was limited internal expertise bridging both worlds. To address this gap, the teams launched a series of joint training sessions aimed at building a shared vocabulary and mutual understanding from the ground up.

But aligning perspectives was only the first hurdle. As the teams began reviewing the models, a new complexity surfaced: data lineage. Tracing where the training data originated—and how it had evolved over time—proved more difficult than anticipated. Much of the data had been pulled from disparate systems, stored across multiple platforms, and often lacked a clear audit trail. Ensuring transparency became a manual, time-intensive process. Technology constraints added another layer of difficulty. The existing systems were not designed to continuously monitor model behaviour. They couldn't flag changes in input data or detect performance drift. New tooling was needed—capable of keeping pace with the dynamic nature of AI models in a high-stakes financial environment.

Amid these technical and procedural challenges, perhaps the most significant shift was cultural. The finance, risk, engineering, and internal audit teams had traditionally operated in silos. This initiative required something more: genuine cross-functional collaboration. Conversations had to be reframed, priorities aligned and trust reestablished. It wasn't just about deploying an AI model—it was about transforming how teams worked together to govern its responsible use.

Several challenges surfaced during the initiative:

- *Limited internal expertise*. At the outset, there was a noticeable lack of shared understanding among the finance, engineering, and internal audit teams regarding the unique risks AI/ML systems posed. Training and joint working sessions were needed to close this gap.
- *Data lineage and availability*. Documenting the full lineage of training and input data proved more complex than anticipated. The decentralized nature of datasets used by the data science team added layers of effort to achieve transparency.
- *Tooling limitations*. Continuous monitoring of model behaviour required investments in new tools and platforms. Existing systems lacked the capabilities to track drift, input anomalies, or changes in performance in real time.
- *Cross-functional collaboration*. The effort relied on effective communication and coordination between teams that typically operated in separate silos. Aligning the priorities and language of the data science and engineering teams with the control mindset of the finance, risk and internal audit teams required sustained effort.

Despite these challenges, the team maintained momentum by emphasizing shared ownership, transparency, and practical alignment between innovation and internal control.

Outcomes

The initiative led to several positive outcomes for Google's finance organization and broader governance efforts:

- *Improved forecast reliability*. With targeted controls in place, the forecasting models delivered more consistent and auditable outputs.
- *Audit readiness*. The introduction of control checkpoints and supporting documentation throughout the model lifecycle allowed both internal and external auditors to evaluate the effectiveness of controls more efficiently. This reduced audit effort and enhanced credibility.
- *Regulatory confidence*. By proactively addressing emerging AI/ML risks, Google demonstrated a clear commitment to regulatory alignment and internal accountability. This effort was positively received during ongoing compliance and governance reviews.
- *Clearer accountability*. The initiative clarified ownership across functional boundaries. Defined responsibilities between the engineering, risk, finance, and audit teams helped embed controls within day-to-day operations rather than treating them as compliance afterthoughts.

The control framework developed through this initiative then served as a model for other AI/ML systems within Google and contributed to broader enterprise-level governance discussions.

Recommendations

Based on the experience of developing AI/ML-aligned controls for financial reporting, the following recommendations emerged for organizations facing similar challenges:

- *Develop a centralized AI/ML risk and control matrix*. Establish a centralized, regularly updated repository that maps AI/ML-specific risks to applicable internal control over financial reporting or SOX requirements. This enabled consistent control design and audit traceability.
- *Implement a formal governance framework*. Introduce structured governance for all high-impact AI/ML models, including defined policies for model documentation, testing, validation, and change management.
- *Conduct a detailed risk assessment*. AI/ML-related risks varied significantly by industry and use case. Performing a thorough risk assessment enabled organizations to identify context-specific threats and select appropriate controls. This risk-based approach ensured that emerging risks introduced by AI/ML models—such as data bias, model drift, or lack of explainability—were proactively addressed in alignment with internal control and compliance frameworks.
- *Adopt automated control monitoring tools*. Leverage automated tools where feasible to support continuous monitoring of model accuracy, data quality, and operational anomalies.
- *Invest in targeted training*. Create and deliver AI/ML-specific control training for the finance, risk, and audit teams. Building shared language and foundational knowledge was essential for effective collaboration.

CONCLUSION

As organizations increased their reliance on AI/ML in financial systems, there became a clear need to evolve internal control practices to keep pace. Google's experience showed that integrating well-designed IT controls into the AI/ML model lifecycle allowed firms to maintain financial integrity while enabling innovation. Rather than treating AI/ML as a compliance challenge, the company approached it as a governance opportunity, demonstrating that it is possible to balance performance, transparency, and accountability in technology-led finance functions. The resulting control model offered a repeatable path for organizations navigating the intersection of AI adoption and regulatory expectation.

EXHIBIT 1: AI/ML MODEL LIFECYCLE STAGES AND CONTROL FOCUS

Lifecycle goals	Defining the objectives and intended use of the model
Business goals	Defining the objectives and intended use of the model
Data sourcing	Identifying and acquiring the data used to train the model
Training	Developing and training the model using the sourced data
Deployment	Implementing the model into the production environment
Monitoring	Continuously monitoring the model's performance and outputs

Note: AI = artificial intelligence; ML = machine learning.
Source: Created by the authors using internal documentation from Google LLC and industry best practices (with sensitive data redacted).

EXHIBIT 2: EXAMPLE CONTROLS FOR AI/ML MODEL LIFECYCLE AND ASSOCIATED SOX RANKINGS AND RISK AREAS

Control type	Example control description	Risk area	SOX ranking
Preventative	AI/ML model's intended use, fairness goals, and key performance metrics documented and approved by finance leadership.	Transparency	Moderate
Preventative	Model design criteria, constraints, and assumptions validated by an independent model validation team.	Model integrity	Moderate
Preventative	A documented and approved framework established for model testing, validation, and tuning.	Governance	Low
Preventative	Raw data sources identified, and their lineage documented and approved prior to model ingestion.	Data integrity	Moderate
Detective	Model performance metrics, including accuracy, precision, and recall, continuously monitored, and deviations investigated.	Model integrity	Moderate
Detective	Input data distributions monitored for significant changes, and alerts triggered for deviations exceeding predefined thresholds.	Data integrity	Moderate
Detective	A comprehensive audit trail of model changes, including retraining, parameter adjustments, and version updates, maintained.	Transparency	Moderate

Note: AI = artificial intelligence; ML = machine learning; SOX = Sarbanes-Oxley Act of 2002.
Source: Created by the authors using internal documentation from Google LLC and industry best practices (with sensitive data redacted).