COSO

Achieving Effective Internal Control Over

GENERATIVE AI

[GenAI]

Authors

Scott Emett
Arizona State University
scottemett@asu.edu

Marc Eulerich
University of Duisburg-Essen
marc.eulerich@uni-due.de

Jason Guthrie*
Ernst & Young Global Limited
jason.guthrie@ey.com

Jason Pikoos
Meta Platforms Inc.
jpikoos@meta.com

David A. Wood
Brigham Young University
davidwood@byu.edu

* The views reflected in this article are the views of the author and do not necessarily reflect the views of Ernst & Young LLP or other members of the global EY organization.

Acknowledgements

We would like to recognize and thank Dr. David Wood for his leadership on this project. Additional thank you goes to the COSO Board, and COSO Board Chair and Executive Director Lucia Wind for providing input, assistance, and valuable feedback in developing this paper. We also thank Professor and Dr. Marc Eulerich, Dr. Scott Emett, Jason Guthrie and Jason Pikoos for their technical input and advice.

This project was commissioned by the Committee of Sponsoring Organizations of the Treadway Commission (COSO), which is dedicated to helping organizations improve performance by developing thought leadership that enhances internal control, risk management, governance, and fraud deterrence.

COSO is a private-sector initiative jointly sponsored and funded by the following organizations:

American Accounting Association (AAA)

American Institute of CPAs (AICPA)

Financial Executives International (FEI)

The Institute of Management Accountants (IMA)

The Institute of Internal Auditors (IIA)

Design and layout: Sergio Analco

Committee of Sponsoring Organizations of the Treadway Commission

coso.org

Contents

Executive summary

Generative AI (GenAI) is arriving in boardrooms and back offices at a speed that traditional governance models never anticipated. Across finance, compliance, operations, and other functions, teams are experimenting with AI copilots, automating reconciliations, and generating analyses at a scale that compresses decision cycles. This acceleration is welcome, but it also introduces risks — additional cyber risks, hallucinations, prompt-injection attacks, opaque reasoning, model drift, and rapid configuration changes — that can threaten the reliability of operations, reporting, and compliance if left unmanaged. This publication builds on COSO's earlier *Realize the Full Potential of Artificial Intelligence* thought piece, extending its enterprise risk management foundation into the realm of GenAI by translating those principles into concrete internal control practices.

This guidance is designed for practitioners responsible for the execution and oversight of AI processes, internal controls, risk management, and assurance, including but not limited to:

- All managers responsible for using AI systems to maintain the quality or efficiency of operations
- Compliance and risk management teams
- Controllers, managers, and financial reporting groups
- IT governance and information security
- Board committees and oversight bodies
- External auditors assessing GenAI-related controls
- Internal audit departments

This report provides a practical, COSO-aligned path to harness GenAI responsibly. Instead of creating a separate governance framework, it adapts the five components and 17 principles from COSO's 2013 Internal Control – Integrated Framework (ICIF) into GenAI-specific practices and organizes common uses into eight capability types spanning the data-to-decision lifecycle. It blends narrative guidance with examples, minimum control expectations, and starter metrics while supporting audit readiness by aligning GenAI controls with established COSO principles so decision makers can test, validate, and rely on them for risk management and control.

What's new here:

- A capability-first taxonomy: Unlike generic AI governance frameworks, we organize GenAI uses into eight distinct capability types (ingestion, transformation, posting, orchestration, judgment, monitoring, knowledge retrieval, and human-AI interaction), each with tailored control requirements that recognize how GenAI risks manifest differently across the data-to-decision lifecycle.
- Audit-ready control mapping: Each of the eight capabilities includes embedded examples, minimum control expectations aligned to all five COSO components, and illustrative metrics designed for both operational monitoring and audit evidence collection, eliminating the gap between governance frameworks and audit requirements.
- Practical implementation artifacts: Beyond conceptual guidance, we provide starter templates including risk assessment matrices, control testing procedures, and metric dashboards that teams can adapt immediately, reducing time-to-implementation from months to weeks.

Introduction

The COSO Internal Control – Integrated Framework (Framework) has remained relevant due to its flexibility across various organizations and its well-defined, comprehensible structure. Its five components — Control Environment, Risk Assessment, Control Activities, Information and Communication, and Monitoring Activities — do not prescribe a particular suite of technology to achieve an organization's objectives; they describe what must be present for an organization to achieve its objectives reliably and reduce the risk level. GenAI doesn't change that foundation; it changes the terrain on which the foundation is built.

GenAI systems ingest structured and unstructured information; generate text, images, and code; and can act on downstream systems through application programming interfaces (APIs), which are standardized ways for software systems to communicate and pass instructions and data. They enable increased efficiencies (e.g., by automating invoice classification and analyzing customer-service interactions) while also expanding analytical possibilities such as synthesizing regulatory updates across multiple jurisdictions or generating forward-looking scenarios. The unique characteristics of GenAI also introduce risks into an organization. For example, GenAI can be confidently wrong; it can be manipulated through crafted prompts; its performance shifts as data changes or a vendor releases an update; and adoption is easy enough that "shadow AI" (unauthorized or ungoverned AI implementations operating outside formal IT oversight) can start outside formal channels.

Our goal is to help you integrate GenAI into your organization by using the Framework's structure to clarify objectives, strengthen control design and execution, and increase rigor in traceability and monitoring. We combine narrative guidance with practical examples to illustrate how controls can be implemented in day-to-day practice.

Figure 1. **COSO-ICIF components**

Scope and boundaries

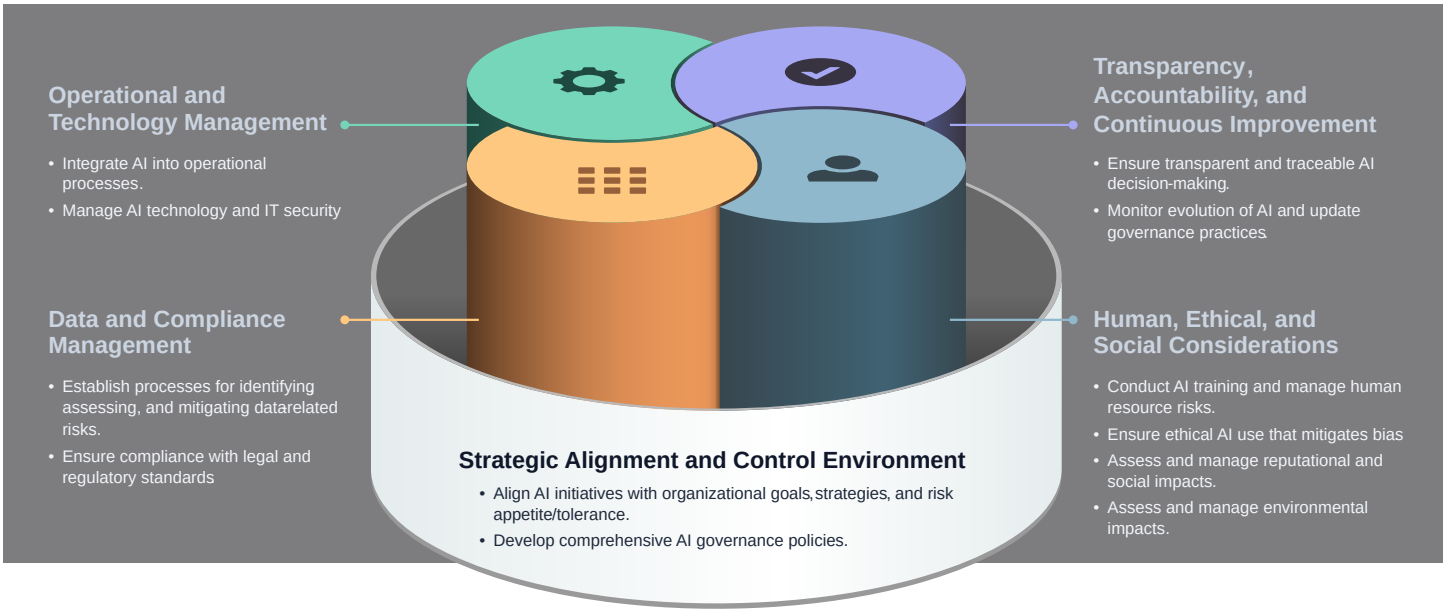
This publication focuses on risks unique to or significantly amplified by GenAI and the controls required to manage them. The principles we discuss frequently apply to other types of AI, but we are not explicitly addressing all forms of AI. We do not repeat established guidance on traditional IT general controls (e.g., network security, non-AI change management, segregation of duties (SoD) in ERP systems), though those remain necessary. For definitions of AI types used throughout this document (and what is excluded), see Appendix A.

We draw examples from financial reporting, compliance, and operations, and the guidance applies broadly across sectors. For organizations seeking a maturity-level view of GenAI adoption, frameworks like the *GenAI Global Framework and Maturity Model* can serve as a complementary reference to the COSO-aligned control approach described here.

The following illustrate GenAI risks addressed in this report:

- **Data quality, source, and completeness.** Inaccurate, unverifiable, or incomplete inputs produce unreliable outputs that are hard to detect once propagated. Incomplete records of the sources complicate auditability and disclosure.
- **Reliability and consistency.** GenAI may produce plausible but factually incorrect or nonsensical information (i.e., hallucinate), undermining accuracy and completeness. Also, changing data, seasonality, model drift or vendor updates can erode reliability and consistency.
- **Explainability and transparency.** Opaque reasoning frustrates validation, testing, and stakeholder confidence.
- **Security and privacy.** The defensive perimeter moves from servers to user interfaces and the underlying data or knowledge sources that GenAI uses. Malicious inputs (e.g., prompt injections) can exfiltrate data, hijack processes, or expose sensitive data.
- **Bias and fairness.** Training data, base models, and retrieval sources can embed bias, creating legal, regulatory, or reputational exposure. Left unchecked, this can also lead to poor decisions that may negatively impact company revenue or drive up unnecessary costs.
- **Third-party and vendor risk.** Any GenAI capability obtained from a vendor —whether hosted, embedded in an application, or deployed on-premises — limits visibility and oversight of training data, model updates, change cadence, data, and underlying control processes.
- **Governance and accountability.** Rapid iteration can outpace existing processes, and items like prompts, thresholds, and retrieval connectors are critical configuration elements that require the same rigor as other controlled system settings.

Figure 2. **GenAI Global Framework and Maturity Model**



AI capability types

Because GenAI manifests in a wide variety of tools, interfaces, and embedded features, looking at it purely by the latest product name or vendor is a moving target. Instead, this publication uses a capability-first lens, focusing on what the AI system actually does — capturing data; transforming it; processing transactions; orchestrating workflows; generating insights; monitoring for anomalies; interpreting organizational and external regulations, requirements, and policies; and interacting with humans.

This capability-first structure builds on COSO's earlier *Achieving Effective Internal Control Over Robotic Process Automation (RPA)* guidance, which aligned RPA governance requirements with the COSO Internal Control – Integrated Framework (ICIF). That publication addressed deterministic automation — rule-based bots that execute predefined tasks — and demonstrated how internal control principles can strengthen automation governance and reliability. GenAI expands this lineage: while RPA follows static logic, GenAI introduces probabilistic reasoning, adaptive learning, and creative generation, extending its influence from task execution to shaping business analyses, judgments, and decisions that can materially affect quality, performance, and profitability. As a result, GenAI requires controls that go beyond execution accuracy to include transparency, validation, and ethical safeguards, ensuring that management can rely on GenAI-supported decisions as part of running the business.

The recurring functions are grouped into eight capability types that follow a logical data-to-decision sequence. This structure makes it easier to see where risks originate, how they can propagate downstream, and where targeted controls can be placed to contain them. The types are intended to be general purpose, applying across industries and processes.

Figure 3. **Eight capability types that follow a logical data-to-decision sequence**

Capability type	Definition	Example(s)
1 Data extraction and ingestion	Capture and interpret raw data from structured and unstructured sources.	Extracts details from incoming customer support emails.
2 Data transformation and integration	Transform raw or unstructured data into usable data by cleaning, normalizing, or combining it.	Apply standardized product categories across e-commerce platforms before analytics are run.
3 Automated transaction processing and reconciliation	Automate high-volume tasks.	Automatically process insurance claims or match supplier invoices to purchase orders
4 Workflow orchestration and autonomous task execution	AI agents coordinate and perform multi-step tasks with minimal human input.	Automatically pull trial balance and subledger data, run reconciliations and analyses, open and assign tasks for follow-up, escalate exceptions, and compile responses into a reconciliation package for review.
5 Judgment, forecasting, and insight generation	Produce forecasts, insights, or draft analyses.	Predict customer demand or generate a market summary.
6 AI-powered monitoring and continuous review	Continuously scan activity for anomalies.	Analyze production sensor data to detect potential equipment failures or quality defects.
7 Knowledge retrieval and summarization	Summarize large volumes of information.	Condense a large volume of new data privacy regulation text for compliance officer.
8 Human–AI collaboration	Augment human capabilities through chat-based interfaces.	Human programmer uses GenAI to write code, which they then review.

Foundational characteristics for GenAI internal control

The capabilities and shortcomings inherent to GenAI have an overall impact on the design and implementation of a system of internal control. The following foundational characteristics can be used to guide how controls should be built or adapted, regardless of the use cases or technology used.

1. GenAI is probabilistic, not deterministic. GenAI outputs are probabilistic and can be confidently wrong. Controls should treat outputs as claims requiring validation, rather than as facts to accept by default.

2. GenAI is dynamic. Models, prompts, and the data they use for retrieval can evolve frequently. Risk assessment, change control, and monitoring should be continuous, or moving toward continuous, to keep pace with these changes.

3. GenAI is easily scalable (for better or worse). Increased automation can scale quality and efficiency, but it can also scale errors and bias. Controls should be designed to prevent small errors from propagating into systemic issues.

4. GenAI has a low barrier to entry. Controls should be designed to govern who can build, deploy and interact with GenAI.

5. GenAI can help govern GenAI. Its analytic and pattern-recognition capabilities can strengthen governance (e.g., through automated multi-model validation that cross-checks outputs across independent models to detect inconsistency, bias, or drift). When properly implemented, GenAI can enhance monitoring, documentation, and validation activities that would otherwise be impractical at scale.

With the capabilities defined, common risks outlined, and foundational characteristics established, we can now integrate them into the Framework in a way that is both comprehensive and practical. The following section includes considerations for each of the Framework's five components of internal control and related COSO principles. This allows us to explore each component in depth, weaving in the associated COSO principles, their effect on the organization, and ideas for practical application. The goal is to make each section a self-contained guide: showing how the component applies to GenAI, which COSO principles are most relevant, how they manifest across capability types, and where targeted controls and metrics can be implemented. A more detailed application of these concepts by capability type is in Appendix B, providing a reference for practitioners who need to design, test, or benchmark specific controls.

Applying COSO to GenAI — Control Environment

A durable control environment sets the tone, defines the boundaries, and establishes accountability for GenAI use across the organization. Within the context of GenAI, the control environment must respond to the technology's accessibility, rapid pace of adoption, and potential to bypass traditional approval channels. Embedding relevant COSO principles ensures that the foundation for responsible AI use is strong and aligned with the organization's objectives.

Principle 1 – Integrity and ethical values
GenAI must operate within clear ethical boundaries. This starts with a GenAI Acceptable Use Policy (AUP) that explains what types of data are prohibited (such as personal information, health data, or other regulated content), sets expectations for avoiding bias, establishes clear limits to use cases that may be unfair or prohibited by law (such as employment decisions or the use of copyrighted materials) and commits the organization to transparency. These values should be evident not just in policy documents, but in everyday review practices and decision-making.

Principle 2 – Oversight responsibility
The board of directors must have visibility into GenAI use and associated risks. This may include formal governance forums, such as a cross-functional responsible AI committee (legal, compliance, IT, and risk), that receive regular reporting on adoption, key risk indicators (KRIs), incidents, and material changes to high-impact use cases, and communicate relevant information to the board. Oversight bodies should have the capacity to challenge assumptions, request independent validation, and direct corrective actions.

Principle 3 – Structure, authority, and responsibility
Clear ownership of AI systems is essential. Each GenAI tool, platform, or organization should have a named owner responsible for its objectives, risk profile, and controls. A responsibility assignment matrix (e.g., RACI) should be developed for key assets such as user prompts, system prompts, retrieval datasets, and transformation rules. Responsibilities should extend to escalation paths for AI-related risks or ethical concerns. Owners must have both the authority to make necessary changes and the accountability for their outcomes.

Principle 4 – Commitment to competence
Assigning ownership without capability invites failure. Employees should receive GenAI awareness and governance training that is appropriate to their role and exposure. For most staff, this may mean understanding acceptable use boundaries and recognizing when to escalate issues, while technical and managerial roles require deeper competence. Engineers and developers should be trained in secure prompting, bias mitigation, and change control, and managers should be equipped to interpret model performance metrics. To establish an AI-ready control environment, organizations may also need to invest in recruiting AI-literate talent within audit, compliance, and risk functions. Training should be continuous and adapted as models, risks, and use cases evolve.

Principle 5 – Accountability
Performance objectives should tie directly to GenAI outcomes where relevant. This includes not only adoption rates, but may also include accuracy, safety, compliance, user adherence to controls, and business growth and optimization. Consequences for misuse or negligent oversight of GenAI systems should be clearly established and communicated. Accountability mechanisms include regular performance reviews, targeted retraining, and escalation for repeated control violations. GenAI can also be used to enhance and improve monitoring against performance objectives.

Common control activities across AI capabilities

- Ownership and boundaries: Assign clear owners for each capability type with defined authority, escalation paths, and documented scope of use.
- Governed configurations: Treat prompts, system prompts, retrieval connectors, and transformation rules as governed configurations with version history, approval workflows, and rollback plans.
- Cultural integration: Embed GenAI governance into the broader control culture, ensuring that innovation is encouraged but guardrails are respected.

Focus areas for specific AI capabilities

While every capability type requires a sound control environment, certain types are especially sensitive to weaknesses at this foundational level:

- Data ingestion and extraction: This is where provenance, data classification, and permissible use boundaries are first established. If controls are weak here, every downstream process inherits compromised or non-compliant data, making remediation difficult and costly. Strong ownership and rules at ingestion prevent risk from propagating.
- Judgment, forecasting, and insight generation: Because these outputs often inform high-impact strategic, business, or compliance decisions, the competence and independence of reviewers is critical. A robust control environment ensures that only qualified personnel approve such outputs and that assumptions, limitations, and contrary information are documented.
- Human–AI collaboration: These tools are often the most accessible and least formally governed entry point for GenAI. Without clear boundaries and training, users may inadvertently input sensitive information, rely on outputs without verification, or share unverified outputs externally. Embedding guardrails here helps maintain organizational integrity and compliance.

Example: Clause extraction assistant

A legal-summary assistant is deployed to extract key clauses from contracts. The legal department sponsors the tool, restricts inputs to approved document classes, and requires all authorized users to complete a pre-release training checklist. New clause types undergo side-by-side pilot testing before being added to production, with results reviewed by the capability owner.

Applying COSO to GenAI — Risk Assessment

Identifying and analyzing risks across an entity's objectives forms the basis for deciding how those risks will be managed. In the context of GenAI, this process must be more dynamic than in traditional environments because models, datasets, prompts, and configurations can change rapidly — sometimes without visible notice. The ability to anticipate, detect, and respond to these changes is essential for maintaining effective control when GenAI is embedded within the control environment.

Principle 6 – Specifies suitable objectives
Every GenAI use case should have clearly defined objectives, boundaries, and success criteria that support the organization's overall goals. Trust or assurance attributes should be measurable design goals for each GenAI use case. Without clarity, risk assessment becomes reactive, chasing issues rather than anticipating them. Objectives should explicitly state acceptable and unacceptable GenAI use cases, their intended users, and any regulatory or contractual requirements it must meet.

Before assessing risks, organizations should also determine whether GenAI is the right tool for the use case. In some cases, deterministic automation or traditional machine learning provides greater reliability and lower risk. Using GenAI where simpler technologies would suffice can introduce unnecessary risk.

Principle 7 – Identifies and analyzes risks
For each use case, assess how GenAI-specific threats such as bias, drift, hallucinations, provenance gaps, prompt injection, and third-party dependencies could prevent objectives from being achieved. This requires considering risk scenarios that combine likelihood and impact. Consider how risks might evolve over time as data sources change, new capabilities are added, vendors release updates, models are retrained, and organizational and external regulations, requirements, and policies change.

Principle 8 – Assesses fraud risk
GenAI introduces novel mechanisms through which fraud can be introduced, including deepfakes, synthetic records, and model manipulation through crafted prompts. These risks can be exacerbated or accelerated through the use of AI agents that introduce authorization risks, excessive agency, and insecure interfaces. Fraud risk assessment should identify how these threats could be exploited internally or externally, and whether existing fraud controls can detect or prevent them. Where necessary, expand preventative programs and detection mechanisms to cover AI-specific schemes..

Principle 9 – Identifies and analyzes significant change
Significant changes in GenAI's model architecture, training data, prompts, retrieval connectors, safety filters, or integration points can materially alter risk profiles. Risk assessment must track these changes and trigger re-evaluation before, not after, they affect operations. This includes vendor-driven changes, which may require specific notification obligations or independent verification, business process changes, inherent changes in model performance (e.g., model drift) and regulatory changes that may affect technology and related policies.

Common control activities across AI capabilities

- Scenario analysis: Ask "What if…" questions for each capability to surface hidden dependencies or edge cases. These scenarios should be documented so they can be used not only for risk management, but also as evidence in audit planning and walkthroughs with internal and external auditors.
- Risk registers that evolve: Maintain living risk registers that update when models, corpuses, or configurations change — not just at annual review cycles.
- Embedded monitoring triggers: Link identified risks to specific KRIs, dashboards, or alerts that will surface early signs of drift, bias, or misuse.

Focus areas for specific AI capabilities

Certain types of AI capabilities warrant heightened attention during risk assessment because the impact of missed or misunderstood risks is particularly severe:

- Data transformation and integration: A small mapping or enrichment error can silently corrupt large datasets, leading to cumulative downstream reporting or compliance failures. Risk assessment should stress-test transformation logic against edge cases and unstructured inputs.
- Automated transaction processing and reconciliation: Misclassification or threshold misalignment can lead to inappropriate action (or inaction) at scale. Risks should be evaluated not only for precision and accuracy but also for impacts on downstream processes.
- Knowledge retrieval and summarization: Incomplete sets of information or misinterpretation of unstructured data can result in inaccurate results. Assess the completeness of coverage and the reliability of AI-generated interpretations before operational reliance.

Example: Forecasting drift

A forecasting model relies on historical sales and macroeconomic indicators. The risk assessment includes drift scenarios tied to macroeconomic shifts, with KRIs that trigger retraining or rollback when variance from actuals exceeds tolerance for two consecutive periods.

Applying COSO to GenAI — Control Activities

Control activities are the actions established through policies and procedures that ensure risk responses are executed effectively. In the context of GenAI, control activities must account for probabilistic outputs, rapid configuration changes, and the possibility of automation operating without human intervention. Decisions about where to place human review gates, how to route exceptions, and how to manage changes are central to this component.

Principle 10 – Selects and develops control activities

GenAI control activities should be designed to mitigate risks to the organization's objectives, which now include GenAI-specific risks such as hallucinations, bias, prompt injection, data leakage, model drift, and over-reliance on GenAI outputs. They should require a level of human corroboration (i.e., human-in-the-loop) proportionate to the risk being addressed, treating GenAI outputs as assertions requiring evidence rather than facts. In areas of significant impact, such as business, legal, or regulatory contexts, segregating GenAI assistance from authoritative decisioning can protect against undue reliance or lack of professional judgment. The selection of control activities to respond to identified risks should consider the effect of potential errors, the speed at which harm can propagate, and how difficult it may be to detect given the changing nature of GenAI systems.

To ensure control activities remain transparent, repeatable, and auditable, organizations should document GenAI-related risk assessments, configurations, control designs, and testing results in proportion to their significance and regulatory exposure. Documentation should be sufficient to demonstrate what was evaluated, who approved it, and when it changed — but not so extensive that it undermines efficiency or scalability. The goal is traceability, not paperwork: to enable validation, learning, and continuous improvement without recreating the burden GenAI is meant to reduce.

Principle 11 – Selects and develops general controls over technology

General controls over technology should be updated to reflect GenAI system lifecycles, model providers, data pipelines, and orchestration layers in addition to traditional applications and databases. Treat GenAI models (and their configurations, fine-tuning artifacts, embeddings, RAG indexes, etc.) as configuration items subject to access control, change management, and other IT operations controls just like any other IT asset. This includes access restrictions, SoD, and documented approvals for changes. Continuous integration/continuous deployment (CI/CD) practices should include independent validation steps (human or automated) for AI-related changes. Where GenAI processes affect areas of significant impact, these general controls should be documented in a way that the AI "bill of materials" contains enough information to support investigations or audits, such as logs of prompts, outputs, system messages, model version, parameters, and plugins used.

This COSO GenAI publication provides a broad, principle-based approach for managing AI-related risk and controls. Recent industry discussions on AI governance and assurance have begun to explore how these same principles apply to internal control over financial reporting (ICFR). These emerging views translate general COSO and AI governance concepts into ICFR-specific considerations such as scoping, reliance determinations, and evidentiary expectations for audits.

To promote consistency with this evolving body of practice, we adopt the following working definition of AI reliance: reliance occurs when management depends on outputs generated by an AI system as part of the evidence supporting an ICFR control's design or operating effectiveness.

- **Reliance example:** AI automatically matches journal entries to supporting documentation, and management reviews and approves the posting using only the AI output.
- **Non-reliance example:** AI suggests matches, but a control owner re-performs the match on all or a sample of items and documents their review and approval.

When reliance exists, AI control patterns should meet the same evidence standards expected for ICFR, including:

- Documented prompt, configuration, and model version
- Clear sampling rationale and exception resolution
- Retention of sufficient evidence to support the design and operation of controls

For non-reliance cases, the guidance in this publication remains relevant, but evidence expectations may be proportionately lower.

Principle 12 – Deploys through policies and procedures
Document how control activities are to be executed, who is responsible, and how they are tested or validated. GenAI use policies should be role- and risk-based when defining allowed tools, prohibited data uploads, and high-risk use cases requiring pre-clearance of documentation expectations (e.g., when to cite sources). These procedures will need to be reviewed more often due to the pace of change with GenAI systems. Depending on the complexity of the GenAI used, personnel may need to be upskilled or retrained so that control activities are being performed by competent personnel capable of understanding the risks inherent to the use of GenAI, identifying errors and taking corrective action, as needed.

Common control activities across AI capabilities

- Validation and hindsight analysis: Test AI performance before and after deployment; periodically retest to confirm ongoing reliability.
- SoD: Separate the ability to configure AI settings from the authority to approve or review outputs.
- Hallucination guardrails: Implement rules that require additional review, require source citations, or block the ability to take action when the confidence of the output falls below acceptable levels.
- Change control discipline: Require documented approvals and evidence for changes to prompts, thresholds, and retrieval corpuses.

Focus areas for specific AI capabilities
Some types require especially robust control activities due to the nature of their outputs or the speed at which errors can scale:

- Data ingestion and extraction: Apply confidence thresholds and require human review for low-confidence extractions and require dual review for use cases before production use.
- Workflow orchestration and autonomous task execution: Simulate and test routing changes before they go live and document expected routing logic so deviations are detectable.
- Judgment, forecasting and insight generation: Require citations for all material outputs and implement capture of contrary information for cases where reviewers disagree; perform hindsight analysis of forecasts against actual results.

Example: Reconciliation auto-posting
A GenAI reconciliation agent is configured to automatically post only when the match confidence exceeds a validated threshold and no policy exceptions are triggered. All other items are routed to a queue with sufficient context for human review. Any change to the posting threshold requires multi-party approval, logged evidence, post-change sampling to confirm accuracy, and regular monitoring for changes in model behavior.

Recent discussions across the assurance and risk management community have identified several leading approaches that can be used to operationalize COSO principles and achieve reliable, auditable AI control performance. The combination of approaches depends on the risks identified, objective of the control, complexity of the AI system, and other controls that may be operating.

1. **Human-in-the-loop (HITL) review:** ranges from full re-performance of AI outputs to risk-based sampling of exceptions
2. **Performance testing:** use of test populations to validate AI accuracy and completeness and to stress-test edge cases
3. **Multi-Model validation:** comparison of outputs across independent AI models or deterministic benchmark algorithms
4. **Data analytics monitoring:** continuous monitoring of AI outputs for anomalies or drift, with thresholds calibrated to risk
5. **Third-party validation:** independent review or certification of AI models and outputs, especially in high-risk scenarios or when relying on third-party AI systems

These methods reflect emerging-industry practice for applying COSO's control principles to AI-enabled processes, supporting both management's assurance activities and auditor expectations for sufficient appropriate evidence.

Applying COSO to GenAI — Information and Communication

Information and communication ensure that relevant, quality information flows to those who need it in a timely manner, both inside and outside the organization. In GenAI contexts, this requires particular attention to source, traceability, and the clear articulation of limitations. Without these, stakeholders may misinterpret AI outputs, over-rely on them, or fail to detect when the outputs are no longer fit for purpose.

Principle 13 – Uses relevant, quality information
Processes using GenAI should capture and store all information to understand, validate, and assess output similarity based on the level of explainability needed for the control activities. This includes prompts, inputs, outputs, source references, model/configuration versions, and any confidence scores – all of which can affect control conclusions. Quality information also means identifying model performance, drift and hallucination metrics, and known limitations so that users can apply judgment when interpreting outputs.

Principle 14 – Communicates internally
Internal communication must ensure that everyone involved in a GenAI-enabled process understands their role, the boundaries of the AI's use, and the procedures for raising concerns. This can include incident notification protocols, alerts for material configuration changes, and periodic updates on model performance or limitations. Messages should be tailored so that operators, reviewers, managers, and governance bodies each receive the right level of detail.

Principle 15 – Communicates externally
When GenAI materially affects customers, partners, regulators, or investors, communication with these external stakeholders about the use, impact, and limitations of GenAI should be clear, accurate, and aligned with organizational and external regulations, requirements, and policy expectations. This may involve disclosures in reports, public statements about limitations, or proactive communication with stakeholders when a significant issue is identified.

Common control activities across AI capabilities
- Source capture: Record where data came from, how it was processed, and by which model configuration, so output similarity can be assessed.
- Centralized repositories: Maintain prompt libraries, retrieval knowledge sources, and model cards in controlled systems with role-based access.
- Output quality KPIs: Define model or output KPIs (e.g., hallucinations, citation coverage, bias) and report them alongside control KPIs.
- Communication protocols: Define who must be informed about incidents, changes, and limitations — and how quickly.

Focus areas for specific AI capabilities
Some types are especially dependent on clear and accurate information flows:
- Data transformation and integration: Communicate any transformation rule changes to all dependent processes before they take effect to avoid silent downstream errors.
- Knowledge retrieval and summarization: Distribute changes to underlying knowledge libraries, policies, or regulatory expectations to all impacted teams in plain language summaries; track acknowledgments where necessary.
- Human–AI collaboration: Display disclaimers or warnings in the interface when outputs have not been verified or may contain sensitive information.

Example: Compliance monitoring digest
A compliance team distributes a weekly summary of GenAI policy violation detections. The summary includes detection metrics, the model/configuration version, data sources used, notable incidents, and any upcoming changes to detection logic.

Applying COSO to GenAI — Monitoring Activities

Monitoring activities evaluate whether the components of internal control are present and functioning. In the context of GenAI, this means verifying not just that controls exist, but that they remain effective on an ongoing basis as models, data, prompts, and vendor capabilities change, often on short notice. Effective monitoring blends continuous observation with periodic deep reviews, ensuring issues are detected early, analyzed thoroughly, and remediated promptly.

Principle 16 – Conducts ongoing and/or separate evaluations

GenAI-enabled processes should be monitored continuously for key performance and risk indicators such as accuracy, precision/recall, coverage, latency, exception volumes, competence, and fairness (where applicable). Given the rate of change in technology, ongoing monitoring activities are critical to identifying and responding to potential issues quickly. This includes regular reviews of AI-generated output to identify anomalies that are not investigated timely or patterns in control exceptions that suggest an issue with the underlying AI model or design of the system.

Some GenAI risks, such as a gradual degradation of the accuracy ("model drift") may be difficult to detect with ongoing monitoring activities alone. Separate evaluations such as periodic model effectiveness audits using historical and hypothetical data, independent challenge sessions, review of the effectiveness of control operators and control design, or simulated adversarial exercises provide a fresh perspective and help uncover issues that ongoing monitoring may miss.

Principle 17 – Evaluates and communicates deficiencies

Just like traditional processes, deficiencies in GenAI processes should be evaluated for severity, root cause, and scope, then communicated to those responsible for corrective action. However, deficiencies caused by the use of GenAI may not only be control gaps but also include safety and reliability failures such as hallucinations in regulated disclosures, biased outcomes, privacy leakage or prompt-injection susceptibility. The root cause analysis patterns for GenAI also need to be updated to include deficiencies stemming from configuration errors, retrieval issues, prompt design, data quality issues, model changes outside change control, or vendor changes.

Communication should extend to governance bodies when deficiencies have material impact, along with an action plan and timeline for remediation. Documenting these evaluations with clear remediation logs, owner sign-offs, and supporting evidence strengthens both ongoing oversight and the organization's ability to demonstrate control effectiveness during internal or external audits.

Looking ahead

Because GenAI systems are probabilistic rather than deterministic, the very definition of a "control failure" may need to evolve. Organizations will increasingly rely on multi-metric AI tolerances rather than relying on a single performance indicator (e.g., pass/fail). This structured approach can be used when evaluating and accepting the performance of an AI system across multiple quality dimensions. For instance, performance may be defined by acceptable error levels or risk thresholds across dimensions such as task accuracy, data leakage tolerance, bias levels, explainability minimums, cybersecurity posture, and model change velocity. Establishing and monitoring these types of tolerance ranges may provide a more realistic view of performance and control effectiveness in dynamic GenAI environments.

Common control activities across AI capabilities

- Combine dashboards for real-time metrics with scheduled deep-dive reviews to validate effectiveness.
- Human-in-the-loop quality reviews for a sample of transactions with use-case-specific rubrics (e.g., accuracy, completeness, tone)
- Establish explicit triggers for retraining, reconfiguration, or rollback based on monitored metrics or incident thresholds.
- Maintain a remediation log that records the issue, root cause analysis, corrective action taken, and follow-up testing results.
- Create an AI control deficiency playbook that includes a mapping of common GenAI failures to standard corrective actions agreed upon ahead of time

Focus areas for specific AI capabilities

Some types require especially vigilant monitoring because of the speed and scale at which errors can propagate or the complexity of evaluating outcomes:

- AI-powered monitoring and continuous review: Ironically, monitoring systems themselves need monitoring to ensure detection logic remains accurate and relevant; recalibration schedules and hindsight analysis are critical here.
- Judgment, forecasting and insight generation: Compare forecasts to actual results regularly, investigate variances, and track performance trends that may indicate model drift or other incremental unintended changes.
- Data ingestion and extraction: Review extraction accuracy when source formats change or new templates are introduced; watch for shifts in confidence score distributions that may signal underlying issues.

Example: Expense monitoring

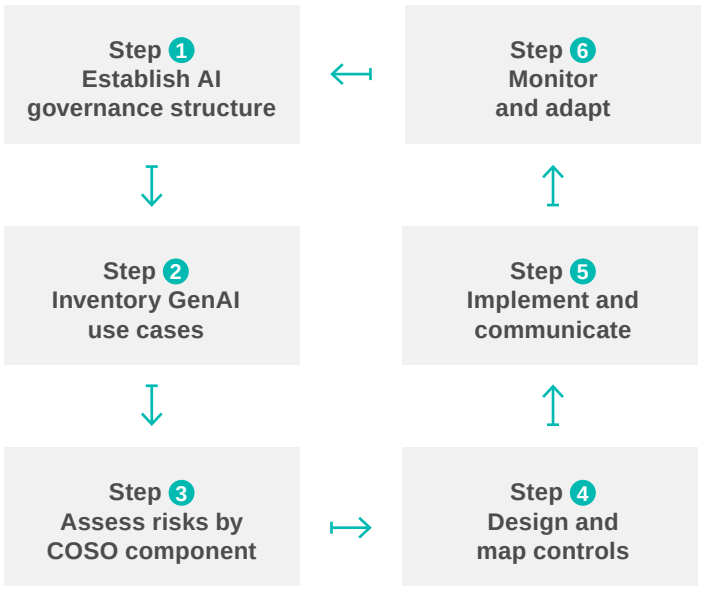
An expense-monitoring model triggers retraining if the monthly precision or recall drops below a defined threshold. Redeployment requires human side-by-side testing against the prior model version, with sign-off from both the control owner and reviewers before changes go live.

Implementation roadmap

The guidance in this publication describes how COSO's five components apply to GenAI and how capability types reveal where risks can originate and propagate based on the use case. The roadmap below turns those concepts into an actionable cycle for embedding GenAI governance into everyday operations. It starts with building a comprehensive inventory of use cases, then moves through assessing risks, designing and implementing controls, and finally monitoring and adapting as technology and circumstances change. This structure ensures that the principles and patterns described earlier are translated into consistent, repeatable practice.

This roadmap is designed to be cyclical: once Step 6 is complete, organizations should return to Step 1 to re-evaluate its governance structure, then inventory and reassess as capabilities, risks, and business priorities evolve.

Figure 4. **Implementation roadmap overview**



Step 1 Establish AI governance structure
Establish an AI governance structure that is appropriate for the size, complexity, and AI usage of the organization. This includes a cross-functional committee that meets regularly to review GenAI initiatives, assess emerging risks, and approve key decisions. Clearly assign responsibilities and document ownership of objectives, risk profiles, and controls.

- **Example:** A cross-functional committee comprised of representatives from legal, compliance, IT, risk management, and business operations that is responsible for setting AI strategy, approving high-impact use cases, reviewing risk assessments, and ensuring alignment with ethical, regulatory, and business objectives.

Step 2 Inventory GenAI use cases
Identify all active and planned GenAI use cases across the organization. This includes periodic scans or surveys to detect "shadow AI" use. Each use case should be classified by capability type, assigned an owner, and documented with key attributes such as objectives, data sources, model used, deployment model, and criticality. Dependencies on other systems or processes, version, and change logs should also be captured.

- **Example:** "Invoice Extraction — Ingestion/Extraction; Owner: AP Ops; Objective: extract header/line item fields; Data Sources: vendor PDFs; Model: third-party (GPT-4o); Deployment: private cloud; Criticality: Moderate."

Step 3 Assess risks by COSO component
Evaluate each use case using the five COSO components and the cross-cutting risk themes described in this report. This assessment should identify GenAI-specific risks such as bias, drift, provenance gaps, prompt injection, third-party dependencies, and SoD conflicts, along with an analysis of their likelihood and impact. Scenario-based risk assessment (e.g., "What if…" exercises) can be used to surface edge cases and dependencies.

- **Example:** For invoice extraction, the risk assessment might examine Personally Identifiable Information (PII) handling, vendor model update cadence, and template drift.

Step 4 Design and map controls
For each identified risk, prioritize the importance of the risk and design preventive, detective, and corrective controls for significant risks. Map each control to its relevant COSO component and associated capability type and determine how its effectiveness will be measured. Metrics and KPIs/KRIs should align with risk appetite and be reviewed at defined intervals.

- **Example:** For invoice extraction, this could include confidence thresholds and human review for new templates, monthly accuracy sampling, and service-level agreements for exception queues.

Step 5 Implement and communicate
Deploy the designed controls and embed them into operational processes. Configure decision gates, dashboards, and automated alerts. Train users and reviewers on the associated procedures and limitations, and publish escalation paths and change histories in accessible repositories. Establish change management protocols for GenAI controls, including transparent stakeholder notification and training on updates.

- **Example:** Disclaimers could be added to AI-generated summaries, and a prompt library and change history could be made available to all relevant teams. Direct feedback mechanisms can be added into the GenAI system user interface.

Step 6 Monitor and adapt
Continuously measure control performance, tracking KPIs and KRIs against established thresholds. Define clear escalation paths for breaches or material changes. Conduct drift and fairness reviews, validate vendor updates before deployment, and report material changes along with their rationale to governance forums.

- **Example:** After a vendor model update, a supply chain team might revalidate demand forecast accuracy against recent actuals, adjust exception thresholds as needed, and provide refresher training to planners on updated escalation procedures.

Integrated case examples

The examples in the main body and Appendix B are intentionally scoped to a single COSO component or capability type, allowing for focused discussion of specific risks and controls. However, in practice, GenAI use cases rarely exist in isolation — a single workflow can draw on multiple capabilities and require coordinated controls across several COSO components.

This section illustrates how multiple capability types and COSO components intersect in real-world scenarios, showing how risks can propagate across domains and how a cohesive control environment — reinforced by clear risk assessment, well-designed control activities, effective information and communication, and active monitoring — manages those risks. The integrated examples help readers see the connections between the individual elements discussed elsewhere in the publication and serve as practical material for training, governance discussions, or cross-functional design sessions.

Case 1 The Disappearing Clause

COSO components in play:

- Control Environment
- Control Activities
- Monitoring Activities

Primary types: data intelligence and extraction, judgment/insight generation, human–AI interaction

A global legal team used a GenAI extractor to identify "termination for convenience" clauses from supplier contracts. Based on testing by the legal team, accuracy was high on clean PDFs but dropped sharply on scanned faxes — a problem that crossed ingestion quality, reviewer competence, and downstream reliance. The ingestion process was updated with a document-type classifier and confidence thresholds, routing low-confidence faxes for manual review. The insight-generation step now required reviewers to confirm extractions before legal decisions were based on them. Quarterly monitoring flagged a vendor optical character recognition (OCR) update that degraded performance; the team executed a rollback within 24 hours. This scenario highlights how ingestion weaknesses, reviewer protocols, and monitoring feedback loops must work together to prevent subtle quality issues from undermining critical legal judgments.

Case 2 The Vanishing Accrual

COSO components in play:

- Risk Assessment
- Control Activities
- Monitoring Activities

Primary types: data transformation and integration, automated transaction processing, judgment/insight generation

A corporate accounting team used a GenAI process to automatically accrue expenses based on historical invoice patterns. When a supplier shifted from monthly to quarterly billing, the model did not detect the change and failed to record the necessary monthly accrual. The omission was caught during the month-end variance analysis, which extended the close process and contributed to an adjustment in the subsequent month. To prevent recurrence, the team implemented "pattern-change" alerts within the data transformation stage, added a controller review of auto-accrual logic during each close cycle, and established variance thresholds that trigger immediate human investigation as part of the close process.

Case 3 The Over-Helpful Copilot

COSO components in play:

- Control Environment
- Information and Communication
- Monitoring Activities

Primary types: Human–AI interaction, regulatory and policy intelligence, AI-powered monitoring and continuous review

A knowledge-worker AI copilot began suggesting draft customer emails that occasionally included internal jargon and non-public roadmap details. This raised issues related to acceptable use policies, external communication protocols, and the need for active monitoring of AI outputs. To address this, the program was updated to add prompt filters and guardrails (e.g., no ability to share drafts externally without human review). Monitoring systems now flag patterns of potentially sensitive output for review, and policy intelligence suggests adjustments to the filters as internal rules evolve. This case shows how human-AI interaction guardrails, policy scanning, and monitoring controls reinforce one another to protect sensitive information in real time.

Conclusion

GenAI's speed, adaptability, and reach mean it will continue to reshape processes, decision-making, and risk landscapes across organizations. These same qualities also introduce distinctive vulnerabilities that require deliberate governance. COSO's Internal Control–Integrated Framework remains a durable foundation for this governance, provided its principles are applied with an understanding of GenAI's unique characteristics.

By combining COSO's five components with a capability-type view of how GenAI is deployed, this report offers both a strategic lens and a practical toolkit. The main body connects components to cross-cutting patterns; Appendix B translates them into capability-specific controls.

The path forward is iterative: inventory and classify use cases, assess risks with a GenAI-aware mindset, design and implement controls tied to both COSO and capabilities, and monitor performance with clear metrics and governance reporting. Done well, this approach transforms GenAI from an emerging risk into a well-governed asset — one that delivers value with confidence, transparency, and accountability.

The sooner organizations embed GenAI governance into their internal control environment, the sooner they can realize its benefits while avoiding the costly risks of uncontrolled adoption.

Appendix A — AI types

Artificial intelligence and automation-enabled solutions include a broad range of technologies. The "type of AI" is intended to provide a common way to describe which AI technologies they are using, to help create consistency in the approach to risk assessments and controls.

For the purposes of this publication we have defined the types of AI technology as follows:

Machine learning (ML)	Enables systems to learn from data without explicit programming, identifying patterns to make predictions or decisions.
Natural language processing (NLP)	Enables computers to understand, interpret, and generate human language.
Computer vision (CV)	Allows computers to "see" and interpret visual information from images and videos.
Generative AI (GenAI)	A subset of ML capable of creating new, original content (e.g., text, images, audio, video) that can be indistinguishable from human-created content. This can extend to models that plan, execute, and adapt multi-step tasks to achieve a defined goal, often by interacting with external tools, data or environments (e.g., AI agents).

For the purpose of this document, "AI system" will be used to define all forms of AI technologies and describe the full system, including all component parts (see scoping below).

The following rule-based automation technologies are not deemed AI and are outside the scope of this document:

Traditional transaction systems and automation	Systems that follow predefined rules and logic to automate tasks, requiring explicit programming for every action and decision. They excel at straightforward, repetitive processes with stable inputs and predictable outcomes.
Robotic process automation (RPA)	RPA involves software bots mimicking human actions to automate repetitive, rule-based tasks by interacting with digital systems at the user interface level. RPA may contain embedded logic, such as if-then statements or decision trees, but it does not learn from data, adapt over time, or generate new outcomes beyond what has been explicitly programmed.[1]

1. For more information on effective controls over RPA, see COSO's *Achieving Effective Control over Robotic Process Automation*, available at https://www.coso.org/rpa-icif.

Appendix B — Detailed COSO mapping by capability type

This appendix provides a deep-dive reference for applying the concepts from the main body to specific GenAI capabilities. Each of the eight capability types from the data-to-decision sequence is presented as a self-contained profile, including example key risks, control considerations across the COSO components, illustrative metrics, and common artifacts (tangible evidence and reference materials that demonstrate a control exists and is functioning as intended).

Use this section when you need to:

- Design or strengthen controls for a specific GenAI capability.
- Identify where and how to monitor GenAI-specific risks.
- Prepare for testing or assurance work tied to a capability.
- Compare your organization's controls against baseline expectations.

While the main body of the report organizes guidance by COSO component to highlight cross-cutting patterns, this appendix flips the view: starting with the capability, it walks through how each COSO component applies in that context. The result is a practical, capability-first checklist that complements the cross-component patterns in the main body.

The examples in this appendix are illustrative, not exhaustive.

Data ingestion and extraction (capture and interpret raw data from structured and unstructured sources)

Capturing data from business processes accurately and securely at the point of entry is critical for the entire GenAI lifecycle. Weaknesses here — such as losing provenance (meaning the failure to capture and retain key details about the origin, source system, or capture method for incoming data) — propagate through downstream processes, undermining reliability and compliance.

Example key risks

- Hallucinated structure or incorrect extraction of fields
- Poor OCR on low-quality scans or non-standard formats
- Hidden Personally Identifiable/Health Information (PII/PHI) in source material
- Poisoned or manipulated inputs
- Provenance loss

Control Environment

- Assign named data owners or custodians for each ingestion channel.
- Publish AUP limits, such as prohibiting PHI in non-compliant systems.
- Train configurators on quality, privacy, and retention requirements.

Risk Assessment

- Evaluate incomplete vs. incorrect capture scenarios.
- Conduct sensitivity analysis for PII/PHI/export control compliance.
- Assess vendor model update impact and template drift risks.

Control Activities

- Configure confidence thresholds with human review for low-scoring fields.
- Require dual review for new templates before production use.
- Pilot changes in a side-by-side test environment.
- Implement schema validation and anomaly alerts.

Information and Communication

- Tag outputs with provenance (file, version, capture method).
- Store field-level confidence scores.
- Publish ingestion standards and verification expectations for downstream teams.

Monitoring Activities

- Maintain accuracy dashboards by document type.
- Conduct periodic rechecks against source material.
- Trigger reviews on vendor/model changes or format drift.

Illustrative metrics

- Field-level accuracy by template type
- Percentage of records below confidence threshold routed to review
- Provenance completeness rate

Artifacts

- Ingestion SOPs
- Approved source list
- Confidence threshold policy
- Retention/encryption plan

Example

A procurement analytics team deploys a GenAI tool to extract and normalize supplier discount terms from contracts. A new vendor template introduced unique phrasing that was misread as "no discount." To prevent similar errors, the team implemented automated flagging for new template types and added a quarterly "template review day" to validate extraction accuracy across all active suppliers.

Data transformation and integration (transform raw or unstructured data into usable data by cleaning, normalizing, or combining it)

Ensuring that data remains accurate, relevant, and compliant during transformation and integration is critical. Errors or bias introduced here can silently corrupt large volumes of downstream outputs, often without immediate detection.

Example key risks

- Silent corruption from mapping errors
- Biased enrichment from training data or rule sets
- Unmapped or dropped values
- Schema drift in source or target systems
- Brittle pipelines sensitive to minor changes

Control Environment

- Appoint transformation logic owners with authority over changes.
- Require cross-functional approval where outputs affect reporting or compliance.

Risk Assessment

- Identify potential for bias or mapping misclassification.
- Evaluate risk of dependency on third-party transformation services.
- Assess impact of schema changes or integration delays.

Control Activities

- Document transformation rules in plain language.
- Maintain automated tests for all rules.
- Require formal approvals for any changes.
- Perform pre-/post-reconciliations between raw and transformed datasets.

Information and Communication

- Maintain a data dictionary with owners and rule lineage.
- Communicate transformation rule changes to all dependent teams before implementation.

Monitoring Activities

- Review error logs and reconciliation variances.
- Conduct periodic integration audits.
- Track time-to-rollback after failed changes.

Illustrative metrics

- Transformation error rate
- Percentage of rules with automated test coverage
- Number of rule changes communicated before deployment

Artifacts

- Rulebook with plain-language descriptions
- Change tickets and approvals
- Lineage diagrams
- Reconciliation templates

Example

A global operations team uses a GenAI process to transform logistics data from multiple carriers into a unified reporting schema. After a schema change at one carrier, fuel surcharge data was dropped from downstream cost reports. A "schema verification" routine was added to run nightly, comparing expected fields against received data and alerting the integration owner if any are missing.

Automated transaction processing and reconciliation (automate high-volume tasks)

Automated posting and reconciliation can greatly improve efficiency but also amplify the impact of configuration errors or misclassifications. Without proper safeguards, errors can cascade through financial systems.

Example key risks

- Misclassification of transactions
- Threshold creep leading to over-automation
- SoD conflicts in configuration and review
- Undetected exceptions

Control Environment

- Define posting authority by role.
- Separate configuration permissions from review responsibilities.
- Document SoD matrices.

Risk Assessment

- Evaluate the materiality of automated transactions.
- Identify potential fraud scenarios from malicious configuration changes.

Control Activities

- Set validated confidence thresholds for auto-posting.
- Route exceptions to review queues with clear context.
- Lock configurations and require multi-party approval for changes.
- Periodically re-perform reconciliations manually.

Information and Communication

- Maintain real-time exception dashboards.
- Attach override rationales to transaction records.
- Keep full audit trails for unattended postings.

Monitoring Activities

- Track error and reversal trends.
- Monitor exception queue volumes and clearance times.
- Conduct post-change stability reviews.

Illustrative metrics

- Percentage of auto-posted transactions within tolerance
- Exception aging time
- Reversal and override rates

Artifacts

- Posting policy
- Threshold register
- Reconciliation runbooks
- SoD matrices

Example

A shared services center uses GenAI-driven reconciliation to match intercompany sales and purchases. During month-end, a misconfigured currency conversion parameter caused multiple matches to be flagged as exceptions, delaying close by two days. The fix included automated validation of currency rates each morning and a "reconciliation health" dashboard for the controller.

Workflow orchestration and dynamic process management (AI agents coordinate and perform multi-step tasks with minimal human input)

AI-driven workflows can adapt processes in real time, improving efficiency but also creating the potential for policy misapplication and hidden SoD conflicts if changes are not well governed.

Example key risks

- Routing rules that create SoD violations
- Policy misapplication due to incorrect logic
- Unapproved changes to routing configurations
- Lack of transparency in prioritization decisions

Control Environment

- Assign process stewards for AI-driven workflows.
- Define approval requirements for autonomous flow changes.

Risk Assessment

- Scenario-test routing under edge cases.
- Assess SoD conflicts introduced by AI decision-making.

Control Activities

- Simulate routing changes before deployment.
- Require multi-party approvals for configuration changes.
- Provide routing rationale in user interfaces.

Information and Communication

- Notify stakeholders when routing logic changes.
- Maintain clear escalation channels for routing disputes.

Monitoring Activities

- Audit routing outcomes for compliance.
- Analyze override and escalation trends.

Illustrative metrics

- Percentage of auto-approved transactions within rules
- Override and escalation rates
- SoD conflict occurrences

Artifacts

- Routing rulebook
- Simulation test results
- Change approvals
- User notifications

Example

An internal audit workflow leverages GenAI to route workpapers to reviewers based on topic expertise. A recent control testing project was accidentally routed to a junior staffer due to incorrect tagging of "expertise" fields. The process steward added a rule requiring a secondary check for high-risk or SOX-related workpapers before routing.

Judgment, forecasting and insight generation (produce forecasts, insights, or draft analyses)

High-judgment GenAI outputs often inform strategic, compliance, or business decisions. The stakes make accuracy, sourcing, and reviewer competence essential.

Example key risks

- Hallucinations or fabricated content
- Use of outdated or incomplete sources
- Automation bias leading to over-reliance on AI outputs
- Ungrounded forecasts or extrapolations

Control Environment

- Require qualified reviewers for all material outputs.
- Formalize decision sign-off thresholds and responsibilities.

Risk Assessment

- Assess corpus completeness and freshness.
- Evaluate potential for bias or selective sourcing.
- Perform sensitivity analysis on key assumptions.

Control Activities

- Require citations for all material outputs.
- Implement dissent capture where reviewers disagree.
- Backtest forecasts and use challenger models.

Information and Communication

- Include sources, model/config version, and known limitations in decision memos.
- Maintain repositories for prompts and source inventories.

Monitoring Activities

- Compare forecasts to actual outcomes.
- Investigate significant variances.
- Conduct independent challenge reviews.

Illustrative metrics

- Forecast mean absolute percentage error (MAPE)
- Hallucination rate from sampling
- Percentage of decisions with reviewer sign-off

Artifacts

- Decision memos
- Prompt libraries
- Source inventories
- Model cards

Example

A finance planning team uses GenAI to model cash flow scenarios. One scenario assumed outdated payment terms from a legacy customer contract, skewing liquidity projections. Controls now require all forecast models to pull payment term data from a current, vetted contract repository before running simulations.

AI-powered monitoring and continuous review (continuously scan activity for anomalies)

GenAI can monitor vast data streams for anomalies, fraud, or compliance breaches. These systems require oversight to ensure detection logic remains accurate and relevant.

Example key risks

- High false positive or false negative rates
- Configuration drift in detection thresholds
- Alert fatigue causing missed incidents
- Overreach into sensitive data without proper controls

Control Environment

- Assign monitoring leads and define scope.
- Maintain investigation runbooks.

Risk Assessment

- Calibrate acceptable false positive/negative rates.
- Evaluate privacy and retention implications.

Control Activities

- Regularly tune thresholds using feedback loops.
- Apply expiration dates to suppression rules.
- Require dual control for detection logic updates.

Information and Communication

- Centralize alerts in a case management tool.
- Provide periodic performance reports to governance bodies.

Monitoring Activities

- Backtest detection accuracy.
- Conduct periodic recalibration.
- Review quality of closed cases.

Illustrative metrics

- Precision and recall rates
- Time-to-detect and time-to-close
- Percentage of rules with assigned owners and review dates

Artifacts

- Monitoring runbooks
- Case taxonomies
- Tuning logs
- Change audit trails

Example

A compliance monitoring system uses GenAI to detect duplicate vendor records in the supplier master file. When new vendors were added in a local currency format, the system failed to detect matches. The monitoring owner added a pre-processing step to normalize all currency fields before evaluation.

Knowledge retrieval and summarization (summarize large volumes of information)

GenAI can synthesize and summarize large volumes of information from diverse sources. Inaccurate retrieval, incomplete coverage, or unverified summaries can misinform decisions, particularly in fast-moving functions like marketing.

Example key risks

- Omitted or stale source content
- Overly narrow retrieval scope leading to incomplete summaries
- Hallucinated facts or unsupported claims
- Misinterpretation of competitor, customer, or market signals

Control Environment

- Assign content owners for marketing intelligence sources.
- Define approved sources and usage boundaries for AI-generated summaries.

Risk Assessment

- Evaluate risk of stale or incomplete data feeds.
- Identify reliance scenarios where mis-summarization could materially affect campaign, pricing, or positioning decisions.

Control Activities

- Require human review for summaries used in strategic or external-facing materials.
- Implement coverage checks and freshness validation for underlying sources.
- Enforce citation requirements so reviewers can trace the basis of GenAI-generated claims.

Information and Communication

- Communicate updates to source libraries and taxonomy changes to all dependent teams.
- Publish guidance on interpreting AI-generated summaries, including known limitations.

Monitoring Activities

- Periodically validate summary accuracy against source material.
- Track drift or degradation in retrieval relevance.
- Conduct sampling reviews of AI-generated content before critical campaign cycles.

Illustrative metrics

- Citation completeness rate
- Coverage ratio across approved marketing intelligence sources
- Percentage of summaries requiring correction during review

Artifacts

- Source inventories and access lists
- Summary review checklists
- Retrieval configuration logs
- Citation policies and reviewer sign-off records

Example

A global marketing analytics team uses GenAI to retrieve and summarize competitor product launches and campaign activity across multiple markets. When one of the syndicated data sources stopped updating due to a vendor API change, the system continued producing summaries that omitted a major competitor's new offering. The team introduced automated freshness checks, added redundancy across data feeds, and required human review for all AI-generated summaries used in campaign planning and executive briefings.

Knowledge retrieval and summarization (summarize large volumes of information)

These tools are often the most accessible GenAI entry point, increasing the risk of data leakage or over-reliance on unverified outputs.

Example key risks

- Prompt injection or manipulation
- Entry of sensitive data into unsecured environments
- Sharing unverified outputs externally
- Automation bias

Control Environment

- Publish acceptable use guidelines.
- Require training on secure prompting and redaction.

Risk Assessment

- Identify leakage and injection risks.
- Define prohibited data entry scenarios.
- Assess high-judgment contexts.

Control Activities

- Apply prompt filters and topic blocks.
- Restrict outputs containing sensitive information.
- Require verification before inclusion in official records.

Information and Communication

- Display disclaimers for unverified outputs.
- Log prompts and outputs.
- Provide help resources for escalation procedures.

Monitoring Activities

- Review logs for risky usage patterns.
- Update filters in response to incidents.
- Target refresher training for repeat issues.

Illustrative metrics

- Percentage of sessions with citations
- Blocked prompt count
- Data leakage incidentsTraining completion rate

Artifacts

- cceptable Use Policy
- Prompt filtering policy
- Logging specifications
- Knowledge base articles

Example

A corporate communications team uses a GenAI assistant to draft quarterly shareholder letters. A draft inadvertently included internal financial forecasts not yet disclosed publicly. To mitigate the risk, the assistant was reconfigured to block insertion of any unpublished financial metrics unless validated by the investor relations team.

About the authors

Scott Emett is an associate professor at Arizona State University. His research examines how producers and consumers of financial disclosures make judgments and decisions, often focusing on how technological disruptions shape those judgments and decisions. He strives to conduct research that offers valuable insights for practitioners in the field, bridging the gap between academic research and practical application. His research has been published in major journals, such as Journal of Accounting and Economics; The Accounting Review; Contemporary Accounting Research; Accounting, Organizations, and Society; Review of Accounting Studies; and Auditing: A Journal of Practice and Theory, among others.

Marc Eulerich, is the Chair for Internal Auditing and the Dean at the Mercator School of Management, University Duisburg-Essen, Germany. He also heads the Center for Internal Auditing Excellence and the Mercator Audit & Artificial Intelligence Research Center (MAARC), both at the same university. He has published over 150 scientific and practitioner articles and books about corporate governance, internal auditing, and strategy. His research is published in numerous national and international journals. Prof. Dr. Eulerich also supports the Global internal audit profession with numerous talks and consulting projects to intensify the relationship between theory and practice..

Jason Guthrie, is a Managing Director within Americas Professional Practice – Auditing (APPAu) based in Cleveland, OH. He leverages his experience in complex accounting and auditing matters to lead the development of digital audit methodology and enablement across the Americas. In his role as Americas Audit Certification Leader, Jason oversees the validation of new audit technology solutions, ensuring they are rigorously assessed and positioned to deliver meaningful impact to audit teams and clients. He represents EY in key industry working groups dedicated to the advancement of analytics and AI in the audit profession. His contributions include service on the PCAOB Data and Technology Task Force and the AICPA Audit Data Analytics Working Group, which authored the AICPA Guide to Audit Data Analytics. Jason is an active alumnus of Brigham Young University, where he earned his Bachelor's and Master's degrees in Accounting with a concentration in Audit and a minor in Information Systems Management. He is a Certified Public Accountant in Ohio.

Jason Pikoos is the Director of Modern Finance at Meta. He works across finance to drive transformaiton, AI adoption and AI governance. Jason was formerly a Managing Partner at Connor Group leading Finance Transformation, Technology and Innovation, including GenAI driven solutions. Jason brings over 20 years of accounting, operational and technology experience, working with high growth and technology companies. He is a leader in helping companies drive operational excellence through processes improvement, technology & automation, data and analytics and effective governance. Jason graduated from the University of Cape Town and spent over 10 years in public accounting prior to joining Connor Group.

David A. Wood is the Glenn D. Ardis Professor of accounting at Brigham Young University. With over 200 publications in respected academic and practitioner journals, monographs, books, and cases, David's research focuses on technology, governance, risk management, and internal controls. His most recent book is titled, "Rewiring Your Mind for AI: How to Think, Work, and Thrive in the Age of Intelligence." His influential work has earned him recognition as one of the 100 most influential people in accounting by Accounting Today. David collaborates with companies of all sizes, accounting firms, and regulators, providing insights and expertise on emerging governance and accounting issues.

About COSO

Originally formed in 1985, COSO is a joint initiative of five private sector organizations and is dedicated to helping organizations improve performance by developing thought leadership that enhances internal control, risk management, governance, and fraud deterrence. COSO's supporting organizations are the American Accounting Association (AAA), the American Institute of Certified Public Accountants (AICPA), Financial Executives International (FEI), the Institute of Management Accountants (IMA), and The Institute of Internal Auditors (IIA).

Notes

Achieving Effective Internal Control Over

GENERATIVE AI

[GenAI]

Committee of Sponsoring Organizations
of the Treadway Commission

coso.org