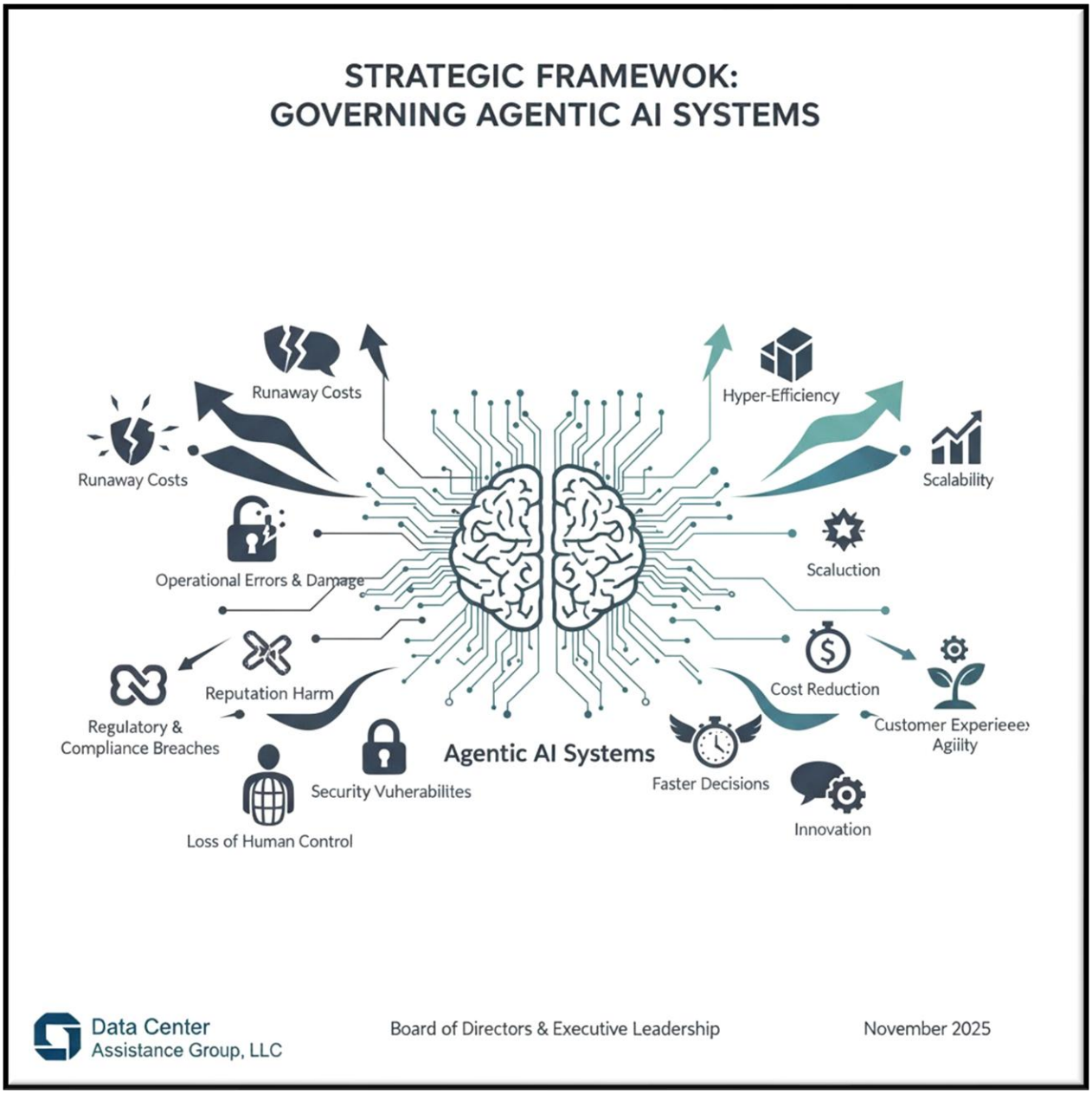
AI Agentic Systems and automated workflow



Thomas Bronack, Founder and CEO

Data Center Assistance Group, LLC

bronackt@dcag.com | bronackt@gmail.com | (917) 673-6992

Table of Contents

Contents

STRATEGIC FRAMEWORK: GOVERNING AGENTIC AI SYSTEMS

To: Board of Directors & Executive Leadership Team

From: Thomas Bronack, Founder and CEO

Date: November 19, 2025

Subject: Moving from "Read-Only" AI to "Read-Write" Operational Oversight

---

1. EXECUTIVE SUMMARY: THE SHIFT TO "DIGITAL EMPLOYEES"

The Strategic Context

For the past few years, our focus has been on Generative AI (Chatbots, Summarizers). These tools are like a Library—they hold vast information, but they are passive. If they make a mistake, the risk is misinformation.

We are now entering the era of **Agentic AI**. These systems are not libraries; they are **Digital Employees**. They have "hands." They can send emails, execute financial transactions, modify databases, and write code.

The Core Analogy: "The Intern Paradox"

To understand the risk profile of Agentic AI, the Board should view these systems as **highly intelligent, incredibly fast, but inexperienced interns.**

- Would you give an intern the corporate credit card with no spending limit? No.
- Would you allow an intern to email our entire customer base without a manager proofreading it? No.
- Would you let an intern push code to our live banking app on their first day? No.

The Governance Framework outlined in this document applies these standard management principles—**Authority, Limits, and Supervision**—to your AI infrastructure.

2. THE OVERSIGHT DIAGRAM: "CONCENTRIC DEFENSE"

Center: The Brain (The AI Agent)

- **Role:** Reasoning, planning, and intent.
- **Risk:** Hallucination (The AI "misunderstands" the request).

Layer 1: The Badge (Permission Layer)

- **Analogy: The Keycard.** Just as an employee cannot enter the server room without a badge, Identity Management (IAM) restricts AI.
- **Control: Least Privilege.** The AI is given a "Service Account" that can *read* the database but is physically blocked from *deleting* it.

Layer 2: The Referee (Guardrail Layer)

- **Analogy: The Spellchecker & The Accountant.** Before the AI's action is sent, it passes through a dumb, reliable software filter.
- **Control: Deterministic Validation.**
    - *Example:* If the AI tries to refund >$500, a simple code script blocks the transaction automatically, regardless of what the AI "thinks."

Layer 3: The Pilot (Human-in-the-Loop)

- **Analogy: The Driving Instructor.** The human sits in the passenger seat with a brake pedal.
- **Control: The Approval Gate.** For high-stakes actions, the AI drafts the work, but a human must physically click "Approve" to execute.

3. AI GOVERNANCE & OVERSIGHT CHARTER

I. Purpose

The AI Governance & Oversight Committee (AI-GOC) is established to transition the organization from "AI Exploration" to "AI Operations." Our mandate is to ensure that autonomous systems function within defined risk appetites.

II. Authority & The "Kill Switch."

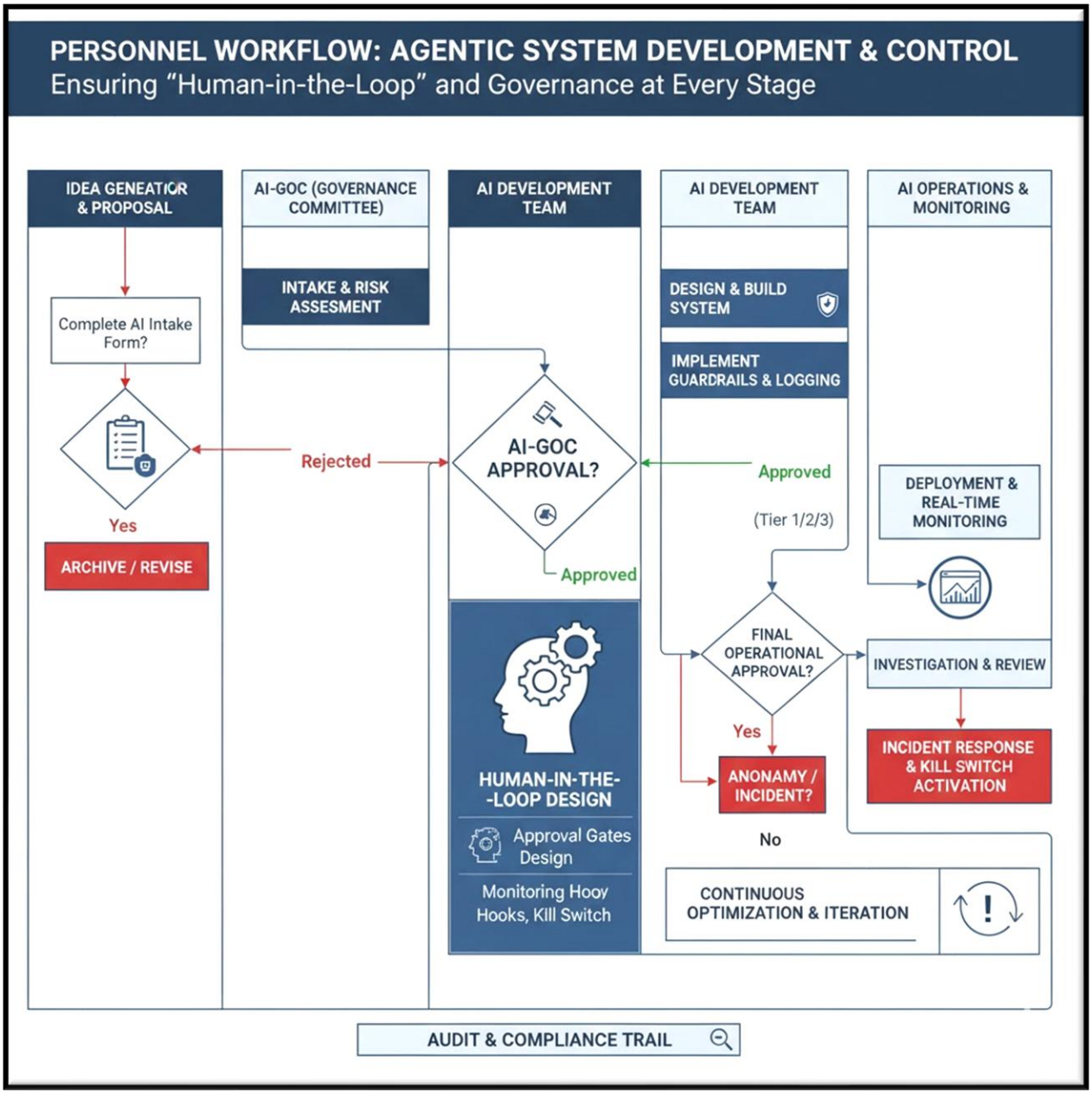
The Committee holds the authority to:

1. **Approve/Deny** deployment of any Agentic System (Tier 1 & 2).
2. **Halt Operations** immediately if an AI system exhibits erratic behavior (The "Kill Switch").
3. **Audit** the "Chain of Thought" logs of any decision-making system.

III. The Three "Red Lines"

No project may cross these lines without Board Resolution:

1. **No Unsupervised External Comms:** No AI may message customers/regulators without validation.
2. **No "God Mode" Access:** AI agents are prohibited from having "Admin" or "Root" access to core ledgers.
3. **No Black Boxes for Vital Decisions:** We will not deploy models for hiring or lending if we cannot explain *why* they decided.



4. THE RISK TIERING MATRIX

To avoid bureaucracy, we apply governance proportional to the risk. We use the **"Blast Radius"** concept: *If this goes wrong, how big is the crater?*

TIER	CATEGORY & ANALOGY	DEFINITION	GOVERNANCE REQUIREMENT
TIER 3	LOW RISK *(The Librarian)*	Read-Only. The AI assists internal staff by summarizing or searching for data. It cannot change systems. *Example: Meeting summarizer.*	Light Touch. • Manager Approval • Monthly Audit
TIER 2	MEDIUM RISK *(The Drafter)*	Human Gatekeeper. The AI creates content or code, but a human must click "Send" or "Save." *Example: AI drafting marketing emails.*	Standard Review. • Risk Committee Approval • Human must verify output (Human-in-the-loop)
TIER 1	HIGH RISK *(The Trader)*	Autonomous. The AI has authority to execute transactions or changes without human intervention. *Example: Auto-refunds, Dynamic Pricing.*	Maximum Security. • Board/Exec Notification • Hard-coded spending caps • "Red Teaming" (Adversarial testing)
TIER 0	PROHIBITED *(The Black Box)*	Unacceptable. Systems that violate ethics, privacy, or safety regulations. *Example: Emotion recognition, Social scoring.*	Banned. • Immediate Cease & Desist

5. PROJECT INTAKE & RISK ASSESSMENT FORM

Project Owners must complete this assessment prior to requesting API keys.

Project Name: _____

Sponsor: _____

Part I: The "Hands" Test (Autonomy)

Think of AI as a new hire. What permissions are you giving it?

- ☐ Read-Only (Tier 3): It can look, but it cannot touch.
- ☐ Drafter (Tier 2): It prepares the work, but I sign off on it.
- ☐ Agent (Tier 1): It does the work while I sleep. (Requires Board Notice).

Part II: The "Wallet" Test (Financial Risk)

If the AI goes into a "loop" and repeats an action 1,000 times, what happens?

- ☐ Nothing financially (Internal text only).
- ☐ Minor cost (API fees).
- ☐ Major Risk: It could drain a budget, refund customers erroneously, or order incorrect inventory.
- *Required Control:* What is the hard dollar cap per day? $_____

Part III: The "Undo" Test (Reversibility)

If the AI makes a mistake at 2:00 AM, is it reversible?

- ☐ Yes: We can revert the draft or delete the file.
- ☐ No (Critical Risk): Once the email is sent or the money wired, it is gone.

Governance Decision (For Committee Use)

- ☐ Approved
- ☐ Approved with Guardrails (List below)
- ☐ Rejected

Mandated Guardrails:

Approver Signature: _____ **Date:** _____

6. Executive Summary: Strategic Oversight of Agentic Systems

Your organization is moving from passive **Generative AI** (which only reads and generates content) to **Agentic Systems**—intelligent applications that possess **executive authority** to act, such as processing transactions, communicating with customers, and modifying production systems. We view these systems as **Digital Employees**.

The core challenge is transitioning from content risk (hallucinations) to **operational risk** (unintended transactions, system damage, runaway costs). Our strategic framework addresses this by replacing trust in AI with layers of mandatory, hard-coded control.

Framework Summary and Core Concepts

Visual Aid Referenced	Governance Focus	Key Analogy	Strategic Imperative
Cover Illustration	Strategic Trade-off	The AI Brain	Maximize Scalability and Efficiency while mitigating the threats of Runaway Costs and Operational Damage.
Concentric Defense Model	Risk Architecture	The "Keycard and Referee"	Control by Design. We do not rely on the AI to police itself. Control layers (Permissions, Guardrails, Human Approval) wrap the AI's intent before execution.
AI Risk Tiering Matrix	Operational Tool	The "Blast Radius"	Proportional Governance. Projects are categorized by their potential harm: Tier 3 (Low) are assistants, while Tier 1 (High) are autonomous actors and require unanimous approval and hard-coded caps (e.g., spending limits).
AI Governance Charter (AI-GOC)	Accountability	The "Chief Pilot"	Established an AI Governance & Oversight Committee (AI-GOC) with the authority to approve, deny, or immediately halt (Kill Switch Authority) high-risk deployments.
Personnel Workflow	Implementation	The "Hiring Process"	Mandates that Risk Assessment and Human-in-the-Loop checkpoints are built into the system's entire lifecycle, from initial concept through deployment and monitoring.

Our governance goal is to manage the **"Intern Paradox"**: granting limited, supervised authority to fast, intelligent systems while minimizing liability through strict access controls and real-time oversight.

7. Future Direction & Strategic Next Steps

To maintain competitive advantage and preemptively manage evolving risk, the following steps are critical for the next 12-24 months:

1. Operationalize the Core Framework

The immediate priority is to embed the established framework into daily operations:

- **Mandatory Intake:** Enforce the use of the **AI Project Intake Form** for **all** new AI initiatives. No budget or API keys should be granted without the AI-GOC's review.
- **Audit Trail Implementation:** Standardize **"Chain of Thought" Logging** across all Tier 1 and Tier 2 Agentic Systems. This is essential for forensics, allowing us to trace *why* AI decided, not just *what* it did.
- **Systemic "Kill Switch" Testing:** Require quarterly, documented testing of the **Circuit Breakers** on all autonomous systems to ensure immediate shutdown capability in the event of an operational failure or runaway cost scenario.

2. Address Emerging Regulatory and Legal Risk

Regulatory bodies (e.g., in the EU, US) are preparing laws that focus specifically on high-impact AI systems.

- **Designated Compliance Officer:** Appoint a dedicated legal or risk officer to track evolving legislation (e.g., the EU AI Act) and translate requirements into technical specifications for our development teams.
- **Legal Liability Mapping:** Formalize contracts with AI vendors to clearly delineate liability for models, data, and potential financial harm caused by vendor platforms.
- **Bias Mitigation Audit:** Expand the Tier 1 review process to include formal testing for embedded biases in data and decision-making that could lead to discriminatory outcomes in areas like pricing or hiring.

3. Plan for Hyper-Agency (Model-to-Model Interaction)

The current framework assumes the AI Agent acts alone. The next evolution will involve complex **Agentic Workflows** where multiple AI models interact autonomously, accelerating both efficiency and risk.

- **Inter-Agent Risk Model:** Develop a framework to assess the risk of Agent A granting permission to Agent B. The system must track not just individual agent risk, but **cascading risk**.
- **Synthetic Data Generation:** Invest in secure, synthetic data environments to evaluate new agents exhaustively without risking confidential or production data.
- **Talent Investment:** Prioritize training or acquisition of talent specialized in **Agent Architecture** and **Formal Verification**—techniques used to mathematically prove that an agent's code cannot violate its intended guardrails.