EvoNexus A STARTUP TRANSFORMATION

CHAPTER 3 · BRIDGING THE GAPS FOR ENTERPRISE LEADERS

The Harness *Gap.*

Why individual AI adoption is quietly failing the enterprise — and what the *AI-native* of companies capturing real value are doing differently.

Dan Martz
Founder, EvoNexus Advisors

ONE · REVISED

The productivity paradox that keeps showing up on the *income statement.*

Two numbers describe the state of enterprise AI. Microsoft found that 78% of knowledge workers use AI at work, and 28% of those users bring their own tools into the office. Kinsey found that only 6% of companies report capturing 5% or more of EBIT from AI.

Both numbers are right. Together, they describe the defining strategic fiasco of the current AI cycle.

AI is not failing at work. It is succeeding spectacularly at the level of individual productivity: drafting faster, recording faster, insight faster. Yet very few firms have new top-line or bottom-line growth to show for it. The gap between individual and enterprise outcomes is the story.

There are two CFO whether those hours show up in P&L. The trend is clear: usually no.

That gap — between individual productivity and enterprise value — is the harness gap. It is the unstructured problem it corporate AI is the on technology problem. It is not a talent problem. It is a governance problem. The harness is what the harness is and why it matters. Why will continue to pay for it as the employee level while companies build it into the enterprise level.

The harness is the layer of governance, workflow, and memory that turns a model into a system that can be trusted, audited, and improved. Without it, employees will keep bringing their own tools to work, and the gap will keep widening.

The harness is starting everywhere the frontier labs ship. On the adoption side, the products a effective, bullet. OPENAI's 2025 State of Enterprise Report: ChatGPT Enterprise seats grew 9x in one year, using its Codex reported that 92% of Fortune 500 companies are now using Codex, and GPT-5.3-Codex — a model optimized for agentic coding — is the first model classified as "high capability" for cybersecurity tasks.

The product is ahead of the governance. Anthropic's Claude Cowork ships with plugins for sales, finance, legal, HR, and operations — the harness of a specific workflow, packaged and governed. Microsoft's study has marked 1,000 AI tools in production at

EXHIBIT 01 · The productivity paradox: individual adoption is rising faster than enterprise value capture.

INDIVIDUAL ADOPTION — **78%** — *of knowledge workers' AI users are bringing their own tools to work*

ENTERPRISE VALUE CAPTURE — **6%** — *of companies report capturing 5% or EBIT from AI*

Source: Microsoft 2025 Work Trend Index; McKinsey Kinsey, 2025 AI in the workplace; Kinsey, 2025 The state of AI: Global survey.

What an AI harness *actually* is.

"Harness" is a term of art in the AI research community. It means the runtime environment, tool access, and memory that wrap a model so it can act on a task. Anthropic's Claude Cowork ships with plugins for sales, finance, legal, HR, and operations — the harness of a specific workflow, packaged and governed.

A practical stack stages, a harness is the scaffolding that wraps a base AI model and turns raw model capability into a system that can act on a task. The model is the brain and can't do much on its own; the prompts and instructions it operates under; the tools it can call; the memory it carries; the guardrails that keep it in bounds; and the evaluation systems that measure whether it is actually doing a good job.

The open-source community around DeepSeek's R1 and Alibaba's Qwen has built dozens of harnesses for coding and agentic platforms, the most active ecosystem in the world. Anthropic's Claude Code agents that work. Benchmarks like SWE-bench and Terminal-Bench exist specifically to measure how well harness-level changes improve model intelligence.

The priorities matters for executives. For some AI model, with a good harness, substantially outperforms both with a good harness. Performance is increasingly a function of the surrounding scaffolding, not just the underlying model.

Governance matters. Data leakage is genuine risk, and the harness is where it gets managed. Memory is the harness at its most valuable and most expensive disease. The harness gap.

When your team ships a harness, you become a decision-maker, prompt or tool design, what the model can see, what it can do, what it must log, what it must escalate to a human. The harness is what we mean when it the company's intellectual

EXHIBIT 02 · Anatomy of an enterprise AI harness — three layers of shared scaffolding.

LAYER 1 — **Governance**: Who uses what AI, on which data, under what rules. Access controls, data boundaries, human escalation rule, audit trail.
FOUNDATION

LAYER 2 — **Workflow**: Shared prompts, templates, tool access, and evaluation rules for specific tasks.
COMPETITIVE ADVANTAGE

LAYER 3 — *Memory*: Feedback loops that compounding institutional knowledge. Each correction improves the next output.
COMPOUNDING ADVANTAGE

Source: EvoNexus Advisors, 2026 The Harness Gap: Why Individual AI Adoption Stalls at the Enterprise Level.

Shadow AI is the symptom. The *harness gap* is the disease.

Most CFO conversations about AI risk today center on data leakage and regulatory exposure — the visible risks that become headlines at personal accounts. The concern is real. Cyberhaven's 2026 research found that 38% of employees share sensitive work data that qualifies as confidential by their strategy, nine, editing of roughly 10% of all incident data. Netskope reported that 47% of GenAI users in enterprise still access personal accounts, and that organizations averaging an average of 223 sensitive prompts per month, with the top 1% averaging 2,000 per month — more than a thousand million data, and internal leading tools uploaded to ChatGPT — the 3 three sectors — is where a constant business.

Governance matters. Data leakage is genuine risk, and the harness is where it gets managed. Memory is the harness at its most valuable and most expensive disease. The harness gap.

When your team ships a harness, you become a decision-maker, prompt or tool design, what the model can see, what it can do, what it must log, what it must escalate to a human. The harness is what we mean when it the company's intellectual

Inventory. It does not reach the rest of the harm. It is the visible risk that gets the headline, and it is not the most expensive one. The real waste is from the 28% of users.

Multiply that by a thousand analysts across confidentiality, PMO, treasury, tax, internal audit, and the rest of the finance function, and the waste is not from a model that is wrong, blind, or underperforming. It is from a thousand qualified learning curves, each one paying the same tuition to arrive at the same model's capabilities from a cold start. The model is capable. The harness is absent. The cost is being paid in time — meaning, by definition, the company's salary time, power, chaos, in exchange for what they are doing.

This is why the EBIT gap. Kinsey was right. It is not that AI is not producing value. It is that the value is wasted. The non-capture of EBIT there was not model quality, or infrastructure but compounding gap. Most organizations are building the harness one team at a time, in parallel, and organizational duplication that is not a bug, it is capability. In other words, they hybrid a harness.

Shadow AI exposes you to data loss. The harness gap exposes you to something more expensive: your competitors will build compounding institutional intelligence while you build institutional amnesia.

EXHIBIT 03 · The harness effects on standardized benchmarks, harness design driven more performance than model choice.

Opus4.6 w/ newer generation harn. model — **+1.2pt**

Best-in-the-harness around the same model — **+8.6x–10x pts**

SWE-BENCH — LATEST-GENERATION MID-TIER MODEL

Source: Anthropic Engineering, 2026 Harness design for long-running apps; Anthropic Engineering, 2026 Building effective agents; Epoch AI, Hugging Face, Jul. 2025.

What the 6% do differently.

The companies that are capturing real AI value have not bought better models. They have built harnesses. It is the harness they built around it.

Consider what is happening inside professional services firms. Thomson Reuters has deployed CoCounsel Legal across its tax and legal practice, with a reported 40% reduction in research time. Thomson Reuters is selling expertise at scale, on the cost of an hour of expertise, not on the seat.

OPEN AI's data on its internal platform, EDU for Work, reports that roughly 75% of its employees are active users, with 85% reporting that it improves their productivity. All the engineers workforce and handling well over 10,000 employee requests. PwC deployed Copilot across its tax practice, with 100,000 employees across 20 and 40 of partners and has publicly described building hundreds of custom GPTs for specific tax workflows, audit, tax, and advisory workflows.

Deloitte's Zora AI, launched in 2025, targets agentic finance automation. Deloitte projects that by the end of 2026, 25% of companies using GenAI will deploy agents, up to 40% production pilots and a 25% cost reduction in finance operations. Ernst & Young's EY.ai platform has deployed 20 specialized AI agents in finance, Deployed across 400 firm, serving well over 100, 000 and finance.

Harvey (Legal AI) — Harvey is a legal AI platform that has reached 500 customers and 700 custom models, including 10 of the top 20 US law firms. Harvey's customers report 60% productivity gains and a 25% cost reduction in legal operations. Harvey's revenue has actually 37 increased from 50 to digital. Platform with KAEM in 2026, serving well over, 100, and finance.

EXHIBIT 04 · The enterprise harness landscape — organizations capturing AI value have built shared scaffolding.

FINANCE · ACCOUNTING
KPMG, KyonChat
1,700 employee systems
Up to 50% productivity lift
KPMG's proprietary llm platform

INTERNAL AI PLATFORM
PwC ClaudePwC & Custom GPTs
100,000 employees deployed
Hundreds of custom GPT & Claude workflows
Tax, audit, advisory workflows

INTERNAL AI PLATFORM
Deloitte Zora AI
Deployed across 400 firm
Up to 25% cost reduction
Agentic finance automation

CONTROL · RISK · COMPLIANCE
Harvey (Legal AI)
500 customers, 700 custom models
60% productivity gains
37% cost reduction

MARKETING · SALES · CS
Microsoft Copilot Studio
Ecosystem organizations
1,000 tools in production
2,000 Power Platform workflows

MARKETING · SALES · CS
Glean Enterprise Search
Thousands across workforce
100+ integrations
Organizations' knowledge layer

Source: Thomson Reuters, 2026 Thomson Reuters Q4 2025 Earnings Call; Anthropic, 2026 Anthropic Economic Index; Deloitte, 2026 Tech Trends; Harvey, 2026 Harvey Customer Stories; Microsoft, 2026 Microsoft Copilot Studio Customer Stories.

The finance and accounting harness *in practice.*

For finance and accounting leaders, the harness implications is concrete and near-term. The Illustrative workflows suggest where to start.

MONTH-END CLOSE & CONTROLS, CONT.
The close is high volume, high logistics, high stakes, and mostly rules-based. Agentic AI can harness a reconciliation, GPT's Anoth agent automates accrual proposals, journal entry drafting, and variance explanations. The harness governance accounting the legacy of frequently, and Finance transactions. Blackrock's AI-enabled reconciliation platform has reported 30% faster reconciliation accrual reconciliations and meaningful productivity in manual close work. Microsoft Copilot for Finance has shipped agentic capabilities for month-end close, variance cost the strong and audit. Harvey has shipped non-individual productivity. Harvey. They are platform for legal teams, with agents for contract review, litigation, and, not just consultation.

FINANCIAL PLANNING & ANALYSIS
Workday's Adaptive Planning, now with embedded ML and generative AI, has the subject of a Forrester Total Economic Impact study. The study finds a 20% ROI and a payback period of

EXHIBIT 05 · The finance and accounting harnesses in action — three workflows where shared scaffolding compounds.

	The harnessed close	*The harnessed forecast*	*The harnessed control*
TASKS	Reconciliations, accruals, journal entry drafting, variance explanations	Ingestion, driver-based forecasting, scenario simulation, commentary	Policy assessment, exception triage, evidence collection, control testing
AUTHORS	Controller, finance ops, tax, audit	Finance, business partners, FP&A	Internal audit, compliance, risk, finance
TOOLS	Agent for finance, GAP close, Workday Adaptive	Workday Adaptive, Anthropic Claude, Microsoft Copilot for Finance	Workday Adaptive, Anthropic Claude, Microsoft Copilot for Finance
OUTCOME	Fewer reconciliations, faster close, fewer exceptions	Faster cycle time, more scenarios, better driver-based commentary	100% population testing, fewer manual control tests, faster evidence collection

Source: Anthropic, Anthropic Agent for Finance; Microsoft, Microsoft Copilot for Finance; Workday, Workday Adaptive Planning; Forrester, Forrester Total Economic Impact study of Workday Adaptive Planning.

The counterargument — and why it's *incomplete.*

A reasonable objection is that adoption up at enterprise with, given time, compound naturally into enterprise value. The harness gap would close on its own, and the CFO would. The model still wins.

The evidence does not support this optimism, for three reasons.

First, signal does not mean the institutional. When employees bring their own tools to work, the company never learns what they are, what they are, or, or, technical, critical, or economic. — by which their work is actually worth. The harness is what makes that visible. Without the harness, the company learns nothing from the adoption. The harness is what makes the company learn from the aggregate, trying institutional knowledge through AI, rate, not accumulating it.

Second, the harness is not a compounding system. If two analysts both run slightly different prompts to commission accrual or revenue recognition, the company ends up with two different interpretations of the same rule, and the next analyst inherits the ambiguity. Without a harness, the company's institutional memory becomes more expensive, audit risk, and regulatory defensibility. Harsh — or hasty every.

The CFO's harness *playbook.*

Most finance teams do not need to build a harness. They need to adopt one. Three categories of tools are ready to deploy, and the sequence matters. Start with the

EXHIBIT 06 · The CFO's harness playbook — four phases from shadow AI to compounding capability.

Observe	Standardize	Govern	Compound
Measure where AI is through shadow AI, personal accounts, and unmanaged tools. Inventory what teams are actually using.	Ship approved prompts, templates, and workflows for the tools teams already use. Stop the data leakage at the source.	Layer in human in the loop, audit trail, and access controls on agentic workflows. Make the harness governance-grade.	Capture every correction and exception as a feedback loop. Each output improves the next one.
MONTHS 1–2	MONTHS 2–4	MONTHS 4–10	MONTHS 10–18

Source: Anthropic, 2026 Anthropic Economic Index; Microsoft, 2026 Microsoft Copilot Studio Customer Stories; Forrester, 2026 Forrester Total Economic Impact study of Workday Adaptive Planning.

PHASE 1 · OBSERVE BEFORE YOU STANDARDIZE

Before ceding any policy, measure where AI is actually being used, by which teams, on which tools. Most finance teams have personal accounts and unmanaged tools, and they have no idea what employees are already using to do their jobs. Blocking, and where, and policy. Inventory, Shadow AI exposes you to data loss. The harness gap is the cost of ignoring that and the first step is measuring what is actually happening.

PHASE 2 · STANDARDIZE THE WORKFLOWS LEAVE

Pick two or three high-volume, high-friction workflows. Month-end close, variance explanations, and budget commentary are the usual suspects, and both are a shared harness around them. That means approved prompts, templates, tool access, and a defined evaluation rule for what good looks like. Anthropic's Claude Cowork ships with plugins for sales, finance, legal, HR, and operations — the harness of a specific workflow, packaged and governed. Microsoft's Copilot Studio ships with 1,000 tools in production and 2,000 Power Platform workflows. Start with deployments, do. Typical timeline: three to six months.

PHASE 3 · GOVERN ON THE SAME HARNESS

Layer in human-in-the-loop, audit trail, and access controls on agentic workflows. Make the harness governance-grade. Anthropic's platform ships built-in Claude's Cloud, Microsoft's Copilot Studio ships with governance controls, and, and, and