THE CIO'S REFERENCE GUIDE

The 12 Levers of Enterprise AI *Adoption.*

Every lever an enterprise has to pull to run AI at scale. Who owns it. What "good" looks like. The single diagnostic question that tells you whether you're pulling it. A reference for the CIO, the CISO, the CFO, and the board.

12	4	10min
LEVERS	CLUSTERS	DIAGNOSTIC

Unmukt Raizada

FOUNDER · TRUSTEVALS

FRAMEWORK EDITION

01 / 2026

Why *twelve.*

Most adoption models are too narrow ("the four pillars of governance") or too broad ("the seventeen dimensions of transformation"). Twelve is where the model stops reducing signal — backwards-derived from CIO, CISO, CFO, and CEO conversations. The four clusters below are ordered left-to-right as most enterprises experience them: visibility first, adoption second, evaluation third, governance fourth.

VISIBILITY

What exists.

I	Discovery & Inventory
II	Usage Depth & Breadth
III	Workflow Integration
IV	ROI & Value Attribution

ADOPTION

Who's using it, how.

V	Training & Enablement
VI	Shadow AI Management
VII	Spend Intelligence

EVALUATION

Is it working.

VIII	Agent Behavior Evaluation
IX	Cross-Org Visibility
X	Change Management

GOVERNANCE

Prove it.

XI	Policy & Governance
XII	Benchmarking & Maturity

LEVER 01

Discovery & *Inventory.*

> *A complete, current registry of every AI system running in or against your organization — vendor tools, embedded AI features inside SaaS you already pay for, and internal agents your teams have built.*

WHO OWNS IT	CIO, with CISO secondary. The single accountable seat is the CIO; the CISO inherits the risk if it goes unmaintained.
WHAT "GOOD" LOOKS LIKE	A registry refreshed automatically — endpoint, network, and SDK signals reconciled — listing every AI tool, every embedded feature, every internal agent, with last-seen timestamps. Most enterprises that look honestly find 150–300+ distinct AI surfaces.
FAILURE MODE	A spreadsheet of approved vendors. Updated quarterly. Already wrong by Tuesday.

DIAGNOSTIC QUESTION

"What AI is running in our organization that we don't know about?"

LEVER 02

Usage Depth & *Breadth.*

> *How deeply AI is being used, not just whether people are logging in. Session depth, feature engagement, and the share of work actually flowing through the AI surface — not the seat count.*

WHO OWNS IT	CIO and the applied AI lead jointly. Marketing-led adoption metrics are the wrong number.
WHAT "GOOD" LOOKS LIKE	Trace-level analysis that distinguishes a user typing one prompt a week from a user driving a workflow daily. Power-user identification is the leading indicator; login counts are the trailing one.
FAILURE MODE	"95% of seats are active." Active is doing one query a month. Login counts always lie.

DIAGNOSTIC QUESTION

"Beyond seat counts — which teams are actually using AI to do the work, and at what depth?"

LEVER 03

Workflow *Integration.*

> *Whether AI is embedded in the critical path of real business processes, or used ad-hoc on the side. Critical-path detection separates substitution ("I asked the bot for ideas") from integration ("the workflow can't complete without it").*

WHO OWNS IT	CIO, applied AI lead, and the line-of-business owner of the workflow. All three signatures or it doesn't ship.
WHAT "GOOD" LOOKS LIKE	A documented map of where AI is on the critical path, where it is a co-pilot, and where it is a detour. The map updates from production traces, not from steering-committee minutes.
FAILURE MODE	"AI is in our workflow." Pressed: which step. Pressed: who owns the handoff. The answers don't exist.

DIAGNOSTIC QUESTION

"In which specific workflows is AI on the critical path — and how would the workflow degrade if you switched it off tomorrow?"

LEVER 04

ROI & Value *Attribution.*

> *A defensible link between AI investment and business outcome — license cost vs measurable productivity, quality, or revenue impact. Cost per outcome, not cost per seat.*

WHO OWNS IT	CFO and CIO jointly. Neither alone produces a credible answer.
WHAT "GOOD" LOOKS LIKE	A cost-per-outcome view that ties AI spend to a unit the business already cares about — tickets resolved, deals closed, hours returned, error rate. The line item the CFO can defend on a board call.
FAILURE MODE	A vendor-supplied "productivity uplift" deck with an asterisk you can't follow. Anchored on a benchmark you didn't run.

DIAGNOSTIC QUESTION

"For every dollar of AI investment, what's the unit of business outcome we can point to — and who signed off on the measurement?"

LEVER 05

Training & *Enablement.*

> *Whether the people in your organization are getting better at using AI over time. Fluency measurement, power-user identification, and the diagnosis of where the enablement gap actually lives — not who attended training.*

WHO OWNS IT	CIO and Chief People Officer jointly. Owned by HR alone, it becomes a learning-management problem; owned by IT alone, it becomes a tool-rollout problem. Both are wrong.
WHAT "GOOD" LOOKS LIKE	A behavior-derived fluency signal — power users identified by what they do in production, not what training they completed. Enablement gaps surfaced from usage data, not surveys.
FAILURE MODE	A completion-rate dashboard for an LMS module. Says nothing about whether anyone got better.

DIAGNOSTIC QUESTION

"Who in our organization is using AI in ways we did not anticipate — and what are they doing that the rest of the org should learn?"

LEVER 06

Shadow AI *Management.*

> *Detection of AI tools running outside approval — and a graduated response: block, allow, or coach. Shadow AI is rarely a single rogue tool; it's a continuous tail of unsanctioned surfaces growing faster than IT can list them.*

WHO OWNS IT	CISO with CIO co-signature. The CISO sets policy; the CIO holds the relationship with the line of business that wanted the tool in the first place.
WHAT "GOOD" LOOKS LIKE	Endpoint, network, and policy signals reconciled into a single view, with three response paths — **block, allow, coach** — applied per tool, per population, per data class. Not a binary firewall.
FAILURE MODE	A blanket block list. Drives use to personal devices. Solves the visibility problem by destroying it.

DIAGNOSTIC QUESTION

"Which unsanctioned AI tools are in active use in our org this week — and for each, are we blocking, allowing, or coaching?"

LEVER 07

Spend *Intelligence.*

> *Whether AI dollars are landing where they create value — or where they don't. License utilization, duplicate tool detection, individual-vs-enterprise pricing, and the infrastructure cost the budget side rarely sees.*

WHO OWNS IT	CFO with CIO partnership. The CFO needs the data; the CIO has the systems that produce it.
WHAT "GOOD" LOOKS LIKE	License utilization per seat per tool, duplicate-tool flags, and a consolidation recommendation refreshed quarterly. Catches the obvious — unused seats, duplicate subscriptions — and the non-obvious — runtime cost growth nobody put on a P&L.
FAILURE MODE	An annual procurement review. Renewal-cycle thinking on a category that moves monthly.

DIAGNOSTIC QUESTION

"For every AI tool we pay for — what share of paid seats are actually used, and what would consolidation save us this quarter?"

LEVER 08

Agent Behavior *Evaluation.*

> *Whether the AI agents you have deployed are actually doing what they should — measured at the level of production traces, not aggregate accuracy claims. Per-stage pass-fail, not end-to-end vibes. Output quality, hallucination, drift, and multi-turn consistency.*

WHO OWNS IT	CIO and CISO jointly, with engineering leadership. The deepest moat — and the lever most enterprises haven't pulled.
WHAT "GOOD" LOOKS LIKE	Production trace evaluation with per-checkpoint pass-fail breakdowns — route, retrieval, payload, execution. Forced-feedback loops where user thumbs-down enters the regression set before the next deploy. Drift detection over weeks, not screenshots.
FAILURE MODE	"Our accuracy is 94%." Asked which pipeline stage. Asked against which baseline. The number is a vibe.

DIAGNOSTIC QUESTION

"For every agent in production — can we show a per-stage pass-fail breakdown, against which baseline, refreshed how often?"

LEVER 09

Cross-Org *Visibility.*

> *A single, defensible view of the entire AI landscape — across vendor tools, embedded AI, internal agents, every department, every subsidiary. The view the board sees and the view the regulator asks for are the same view.*

WHO OWNS IT	CIO. The board sponsor reads this lever's output. There is no shared owner.
WHAT "GOOD" LOOKS LIKE	One pane of glass: every AI surface, every department, every status indicator on each of the twelve levers. Department-vs-department comparisons. The output is reviewable by a board sponsor without translation.
FAILURE MODE	A list of subscriptions, exported to PDF, called "AI strategy." Twelve different teams each maintaining their own dashboard.

DIAGNOSTIC QUESTION

"Could we put a one-page AI landscape report in front of the board this Friday — with a status indicator on each of the twelve levers?"

LEVER 10

Change *Management.*

> *Whether the organization itself — roles, processes, decisions — is being redesigned to use AI well. Not the rollout plan; the structural work. Most failed AI programs are change-management failures dressed as technology failures.*

WHO OWNS IT	CIO, Chief People Officer, and a named change leader. Without all three, this lever is a slide deck.
WHAT "GOOD" LOOKS LIKE	A documented set of role redesigns and process redesigns tied to specific AI capabilities — with measurable adoption curves per redesigned workflow. Decisions captured in durable artifacts, not relitigated quarterly.
FAILURE MODE	Redesign loops. Eight strategy documents. None of them touched the actual workflow.

DIAGNOSTIC QUESTION

"Which roles and processes have we redesigned because of AI — and where is the durable record that says so?"

LEVER 11

Policy & *Governance.*

> *The rules that govern AI use — encoded, enforced, and continuously verified. Policy-as-code, per-use-case baselines, and the recognition that frameworks (NIST, ISO 42001, EU AI Act) tell you what to track but not what "good enough" looks like.*

WHO OWNS IT	CIO, CISO, and Chief Compliance Officer. Three signatures on the baseline; one accountable seat for enforcement.
WHAT "GOOD" LOOKS LIKE	Policy expressed as runnable code, evaluated against per-use-case baselines (a bias metric of 0.12 isn't "compliant" or not — it depends on the use case). Continuous evidence generation, not periodic checklists.
FAILURE MODE	A 47-page policy PDF. Reviewed annually. Cannot tell you if it was violated this morning.

DIAGNOSTIC QUESTION

"What is our defined 'good enough' threshold for each AI use case — and how would we know, this morning, if it had been crossed?"

LEVER 12

Benchmarking & *Maturity.*

> *Where you sit on a real maturity curve — and where your peers sit. Not a vendor-supplied trophy chart; a model with rungs the organization can climb. Most enterprises believe they're at Stage 5 while sitting at Stage 1.*

WHO OWNS IT	CIO, with the board sponsor as the consumer. The maturity score is the artifact the board uses to set next year's investment.
WHAT "GOOD" LOOKS LIKE	A maturity assessment scored against an industry-grounded model — with cross-customer baselines that show where you stand against peers, refreshed at least quarterly. The rung the org thinks it's on, and the rung the data says it's on, are the same rung.
FAILURE MODE	A self-assessment survey filled in by the team that hopes to get budget. Always reads "Stage 4."

DIAGNOSTIC QUESTION

"On a maturity model with rungs — which rung are we actually on, and which rung do our peers sit on?"

How to use this *framework.*

The point of the model isn't a scorecard. It's to let a CIO look at the organization and know, in under ten minutes, which levers are being pulled well, which aren't, and which haven't been thought about. Three actions, one per audience.

FOR THE CIO

Walk into each function (IT, Engineering, Ops, Finance, Compliance) with one question: which of these twelve do you own, partly own, or not own. The gaps are the most interesting thing you'll find this quarter.

FOR THE BOARD SPONSOR

Ask for a one-page AI landscape report with a status indicator on each of the twelve levers. If the team can't produce it, that's the conversation to have — not "do we have an AI strategy."

FOR THE PRACTITIONER

Stress-test vendor claims. Any tool that claims "full AI governance coverage" should be able to name which of the twelve levers it pulls, to what depth, and by what method. If they dodge, keep shopping.

Walk us through your *twelve* — we'll show you the gaps in 30 minutes.

CAL.COM/TRUSTEVALS →