A REFERENCE & SELF-ASSESSMENT INSTRUMENT

The AI-Native Engineering Maturity Model

Ten dimensions for scaling AI — and the decisions that still have to be human.

Ed Schaefer
AI-NATIVE TRANSFORMATION ADVISOR
agile-ideation.com · @agileideation

v1.0 / 2026

WHAT THIS IS

A map for scaling AI — and a way to see where you stand

Most organizations adopt AI tool by tool. This model looks at the system around the tools: how AI is governed, measured, secured, architected, and learned. It gives a leadership team a shared language to have an honest conversation: where are we today, and where do we want to go next?

The shape. Ten dimensions, four stages each. Every dimension follows the same arc — AI use going from invisible, to visible, to codified, to embedded — and the further along a dimension gets, the more of the mechanical work AI takes over. The real question each one asks is where, as that happens, a person still has to own the decision. That's what the model is for: seeing how far AI has come in each area, and where human judgment still belongs.

STAGE	WHAT IT MEANS	WHERE AI USE LIVES
1 · Shadow	Individuals or teams use AI without org awareness, policy, or measurement.	*Invisible* *off the books*
2 · Sanctioned	The org acknowledges AI use and permits it with basic boundaries and an acceptable-use policy.	*Visible* *loosely controlled*
3 · Standardized	Documented standards and repeatable practices, codified across teams.	*Codified* *same shape org-wide*
4 · Integrated	Built into the operating model; continuously measured, governed, and improved.	*Embedded* *how work gets done*

How to use it. Read each dimension and place your org on the four-stage scale. Every dimension lists a few concrete signals to help you tell which stage you're in. Place yourself honestly, then ask a few other people where they'd put the org — the most useful conversation is usually in the disagreements. Focus on the one or two dimensions that matter most to you and your organization right now. Once you've placed yourself, the stage above is your next move, and each dimension's question is a good way into the conversation.

THE TEN DIMENSIONS

At a glance

Each dimension measures one part of the system around the tools, and the last column names the core shift it's about. How far you take each one is a deliberate choice — for some, reaching Integrated is the goal; for others, a solid Standardized is enough.

#	DIMENSION	WHAT IT MEASURES	THE CORE SHIFT
1	**Governance & Risk**	Policy, boundaries, and review for AI use.	*From what we ban to what we enable.*
2	**Agent Autonomy**	How much an agent does on its own, and who decides.	*From blanket rules to earned trust.*
3	**Operating Model**	Roles, decision rights, and review gates around AI.	*Give AI the same role clarity you give people.*
4	**SDLC Integration**	How deliberately AI works across the lifecycle.	*AI works at every stage; the gates are yours to place.*
5	**Leadership**	Whether leaders make non-naive AI calls.	*From watching the work to judging it well.*
6	**Tooling**	How tools are selected, governed, and cost-controlled.	*From invisible spend to cost you can stop.*
7	**Metrics & Measurement**	How impact is measured — activity vs. outcome.	*Knowing what changed, and by how much.*
8	**Security & Data Handling**	How sensitive data is protected through AI tools.	*Seeing the data before it leaves.*
9	**Architecture**	How AI integration survives provider and model change.	*Build for the swap you know is coming.*
10	**Team Capability & Culture**	How fluency, learning, and safety spread on a team.	*From one fluent person to a team that learns.*

DIMENSION 1 OF 10

Governance & Risk

How an organization sets policy, boundaries, and review for AI use — and treats the resulting risk.

STAGE 1 Shadow	AI is used across the org but invisible to leadership. No acceptable-use policy, prompt-hygiene standards, or data rules. Individuals decide what goes into prompts — some of it sensitive. The risk is real, and no one owns it.
STAGE 2 Sanctioned	AI is acknowledged and permitted with basic boundaries. A one-page acceptable-use policy says what data can and can't go in, which tools are approved, and where to escalate. Compliance has been consulted; people aren't guessing.
STAGE 3 Standardized	Governance is codified across teams: documented prompt-hygiene standards, reusable role-spec prompts in version control, and risk-tiered review gates. New use cases go through a defined risk-assessment step before they ship.
STAGE 4 Integrated	Governance is built into how the org operates. Risk monitoring is continuous, and each accepted risk has a named owner. Compliance, engineering, and security stay in one conversation. Governance is used to tell teams what's safe to ship faster — though keeping it current takes real, ongoing effort.

PLACE YOUR ORG — circle where you are

- ○ **1 · Shadow** — no policy or rules; AI use is invisible to leadership.
- ○ **2 · Sanctioned** — someone can point to a written acceptable-use policy.
- ○ **3 · Standardized** — prompt standards and review gates are documented, not just understood.
- ○ **4 · Integrated** — every high-risk use case has a named risk owner.

ASK YOURSELF

Where does your governance actually live today?

AI can draft the policy, surface the risky prompts, and model the odds. Owning the call — and answering for it later — is the part that can't be handed off.

THE MOVE

Apply the discipline to the decisions.

Deliberate calls on the specific ones, plus a standing job to keep the standards current.

DIMENSION 2 OF 10

Agent Autonomy

How much an AI agent is allowed to do on its own — and how that latitude is granted and changed.

STAGE 1 Shadow	No model for what agents can do. Individuals make ad-hoc calls about what agents read, write, or trigger. Some agents are unbounded; others are needlessly boxed in.
STAGE 2 Sanctioned	Agents are held to advise-and-recommend by default. Anything that writes code, deploys, or triggers an action needs a person to approve it. The line is blunt, but it's real.
STAGE 3 Standardized	A tiered trust model — a five-tier agent-trust ladder — defines what each tier can read, write, run, deploy, and approve. New agent classes go through a known promotion path, and the review gate matches the tier.
STAGE 4 Integrated	Autonomy is adjusted by track record: an agent earns more latitude by performing, and loses it when it doesn't. Where a human stays in the loop is decided per use case by what's at stake, not set once for everything.

PLACE YOUR ORG — circle where you are

- ○ **1 · Shadow** — no shared model for what agents may do on their own.
- ○ **2 · Sanctioned** — agents can't write, deploy, or trigger without explicit approval.
- ○ **3 · Standardized** — you can name the tier a given agent sits in and what it may do there.
- ○ **4 · Integrated** — an agent's latitude has actually changed — up or down — based on its record.

ASK YOURSELF

How far does your AI go on its own today?

AI can run inside its tier all day. Deciding which tier it has earned — and when to pull that back — stays a human call.

THE MOVE

Build the trust ladder.

Define the tiers, set the promotion and demotion criteria, and place the human checkpoint by what's at stake.

DIMENSION 3 OF 10

Operating Model

How roles, decision rights, and human review gates are arranged around AI work.

STAGE 1 Shadow	No defined operating model. Who decides, reviews, and owns is worked out case by case. Some decisions get over-engineered; others fall through the cracks.
STAGE 2 Sanctioned	Basic roles are named — who approves a tool, who reviews high-risk uses. Some decision rights are clear; human-in-the-loop placement is informal but acknowledged.
STAGE 3 Standardized	Decision rights are documented across use cases. Operator, reviewer, and approver roles are clear. Where a human review gate belongs is treated as a design decision, since different work needs different gates.
STAGE 4 Integrated	The operating model treats AI as a participant you design for: it gets a role spec, explicit decision rights, and a working agreement for what it's allowed to do. Those evolve as the work changes.

PLACE YOUR ORG — circle where you are

- ○ **1 · Shadow** — who decides, reviews, and owns AI work is worked out case by case.
- ○ **2 · Sanctioned** — someone can say who approves a new AI tool and who reviews high-risk output.
- ○ **3 · Standardized** — operator, reviewer, and approver roles are written down, not improvised.
- ○ **4 · Integrated** — AI has a role spec and decision rights the way a team member would.

ASK YOURSELF

Who's accountable when the AI gets it wrong?

AI can do the work. Who runs it, who checks it, and who owns the outcome is a design decision a person has to make.

THE MOVE

Give AI a real role.

A defined role, clear decision rights, and a gate sized to the stakes — the same clarity you'd give a person.

DIMENSION 4 OF 10

SDLC Integration

How deliberately AI participates across the software lifecycle, stage by stage.

STAGE 1 Shadow	AI is used ad hoc at individual stages — one team for code generation, another for tests, nobody coordinating. SDLC documentation doesn't mention AI; leadership can't see it.
STAGE 2 Sanctioned	AI is explicitly part of some stages, usually coding and sometimes review. Planning, testing, and operations are catching up. The picture is uneven.
STAGE 3 Standardized	AI participates across the relevant stages with a defined human review gate per stage. Documented practices say what AI does and where humans stay in. New integrations follow a known pattern.
STAGE 4 Integrated	AI is designed into the lifecycle, with review gates that move as capability and risk change. Throughput goes up — and review becomes the new constraint, so reviewer capacity and the gates get managed on purpose rather than assumed.

PLACE YOUR ORG — circle where you are

- ○ **1 · Shadow** — AI shows up at scattered stages, with no coordination or visibility.
- ○ **2 · Sanctioned** — AI's role in coding or review is written into how you work, not just happening.
- ○ **3 · Standardized** — each stage AI touches has a defined human review gate.
- ○ **4 · Integrated** — you manage reviewer capacity as a real constraint, not an afterthought.

ASK YOURSELF

Where does AI actually touch your delivery?

AI can take a pass at every stage. Where you put the review gates — and how far you trust each one — is yours to decide.

THE MOVE

Gate by what a miss costs.

Map where AI touches delivery, put a gate where a mistake would hurt, and watch review load as throughput climbs.

DIMENSION 5 OF 10

Leadership

Whether leaders make non-naive calls about AI — informed by enough firsthand context to judge it well.

STAGE 1 Shadow	Leaders know AI exists, but most haven't taken a formal position. A few use it; many don't. Strategy doesn't mention AI; where it shows up, it's off the books.
STAGE 2 Sanctioned	Leadership has named AI as permitted. Strategic intent is stated at a high level — 'we'll use AI where it makes sense' — true and still vague. Leaders are on record but speak about the work from outside it.
STAGE 3 Standardized	Leaders have enough firsthand exposure to make non-naive calls — where to trust AI, where to gate it, what to fund. AI shows up in strategy with measurable goals, and leadership adjusts as adoption data comes in. As the tools speed up, the leadership job grows: intent, review gates, and the safety for people to be honest about what's working.
STAGE 4 Integrated	AI decisions are consistently well-judged, because the people making them have hands-on context and use it to coach others. That experience shapes how the org sets strategy, runs operations, and develops people. Firsthand use is the engine; the maturity is in the quality of the calls it produces.

PLACE YOUR ORG — circle where you are

- ○ **1 · Shadow** — most leaders haven't taken a real position on AI.
- ○ **2 · Sanctioned** — leaders have taken an actual position on AI, in writing, in the strategy.
- ○ **3 · Standardized** — a leader can describe a real AI trade-off they made, not just one they approved.
- ○ **4 · Integrated** — AI decisions are consistently well-judged, and leaders coach others through them.

ASK YOURSELF

Am I leading AI from inside the work, or from the sidelines?

AI can draft the memo and surface the numbers. What's worth doing, who to trust with what, and where judgment has to stay human — that's the job.

THE MOVE

Judge the calls, not the keystrokes.

Get enough hands-on context to make non-naive calls, then hold leaders to the quality of those calls.

DIMENSION 6 OF 10

Tooling

How AI tools are selected, governed, and cost-controlled.

STAGE 1 Shadow	Everyone picks their own tools. The same model is paid for three ways across three teams; spend has no owner and surfaces after the fact. A workload can recompute the same paid result every run and burn budget unnoticed.
STAGE 2 Sanctioned	An approved tool list exists. Procurement tracks basic cost; SSO covers the common tools. Exceptions still happen but are now visible. The org knows total spend but not per-workload cost.
STAGE 3 Standardized	A tool catalog with approval status, per-team cost allocation, and vendor governance. Cost optimization is practiced — cheap-model routing, prompt caching. Guardrails live in code: separate keys per workload, build checks that fail on un-saved paid calls.
STAGE 4 Integrated	Tool and model choices are reviewed against the work as it changes. The org can tie a cost to the workload that caused it and say what value came back. The multi-vendor setup is deliberate, and the cost instrumentation runs where the work runs.

PLACE YOUR ORG — circle where you are

- ○ **1 · Shadow** — everyone picks their own tools and the spend has no owner.
- ○ **2 · Sanctioned** — there's an approved tool list and someone owns the spend.
- ○ **3 · Standardized** — you can allocate AI cost to a team, and guardrails live in code, not a wiki.
- ○ **4 · Integrated** — you can attribute a cost to the specific workload that caused it.

ASK YOURSELF

Would anything actually stop a runaway AI workload before the bill arrives?

AI can route the work, cache what's stable, and log every call's cost. Which spend is worth it — and where the hard ceiling goes — is a human call.

THE MOVE

Put the guardrails in the code.

Persist paid results, isolate the blast radius, and fail the build on a silent-cost bug.

DIMENSION 7 OF 10

Metrics & Measurement

How AI's impact is measured — activity versus outcome.

STAGE 1 Shadow	No AI-specific metrics, so investment runs on perception. Some teams claim big gains, others none, and neither can be checked. With no baseline, no one can tell later whether anything moved.
STAGE 2 Sanctioned	The first numbers count activity — active users, calls, tokens — and spend shows up at the invoice. Real progress, and a trap: volume starts to look like value. The productivity debate is still settled by anecdote.
STAGE 3 Standardized	Metrics shift to quality, flow, and cost per outcome. Delivery measures are recomputed for non-deterministic work — a clean merge means it cleared its review and eval gates. Cost becomes per-user and per-team with ceilings.
STAGE 4 Integrated	Measurement runs as a loop. Outcome metrics tie to business value, and AI performance is tracked against a baseline, so the question is always 'better than what, and by how much.' A jump in usage gets questioned before it gets celebrated.

PLACE YOUR ORG circle where you are

- ○ **1 · Shadow** there are no AI-specific metrics; it runs on perception.
- ○ **2 · Sanctioned** you track any AI number at all, even just usage.
- ○ **3 · Standardized** you measure cost per outcome, not just activity, against a ceiling.
- ○ **4 · Integrated** you compare AI performance to a baseline and act on the gap.

ASK YOURSELF

How do I know whether our AI investment is actually working?

AI can build the dashboards. What a number means, and what to change because of it, is the human part.

THE MOVE

Measure outcome against a baseline.

Set the baseline, track outcome against it, and let the result drive a real operating decision.

DIMENSION 8 OF 10

Security & Data Handling

How sensitive data is protected as it flows through AI tools.

STAGE 1 Shadow	No AI-specific data policies. Sensitive material — customer information, financials, your own IP — may be flowing into prompts, and no one would know. Individuals decide what's safe to paste. The exposure is real and invisible.
STAGE 2 Sanctioned	Basic data classification reaches AI tools. Sensitive categories are named with a do-not-put-into-AI line; SSO and paid accounts cover common cases. Keys and scopes are separated per workload so one exposure stays contained.
STAGE 3 Standardized	Classification is documented and mapped to use cases — which providers may touch what, what's encrypted. Compliance requirements are written down. An audit trail records what AI saw and produced; secure patterns are codified into templates.
STAGE 4 Integrated	AI data security is woven into the overall posture and watched continuously. Controls live in code — a build check can catch a path that sends sensitive data to a provider without recording it, so most unsafe versions get stopped before they ship. Security is used to help teams move faster.

PLACE YOUR ORG circle where you are

- ○ **1 · Shadow** no data rules; sensitive data may be going into prompts unseen.
- ○ **2 · Sanctioned** there's a named do-not-put-into-AI data line people actually know.
- ○ **3 · Standardized** data classes are mapped to providers, and there's an audit trail of what AI saw.
- ○ **4 · Integrated** a control in code catches sensitive data heading to a provider unrecorded.

ASK YOURSELF

Could you say what your AI tools have already seen?

AI can classify the data, enforce the rules in code, and log every exposure. What an exposure is worth, and what the org can live with, is a judgment a person owns.

THE MOVE

Design for visibility first.

Map data to providers, record what AI touches, and put the check in code so the risky path is caught early.

DIMENSION 9 OF 10

Architecture

How AI is integrated technically — and how well it survives provider and model change.

STAGE 1 Shadow	AI is bolted onto whatever exists. A model call is wired in where it fits that afternoon; the provider is hardcoded, the cost-and-latency choice made once and forgotten. Nothing is reusable, and a provider's bad hour breaks the integration.
STAGE 2 Sanctioned	Patterns start to repeat. A couple of teams settle on the same chat or retrieval shape and one writes it down. Architecture reviews ask about AI but stay shallow. The model is still 'whatever SDK we use,' so swapping providers means rewriting consumers.
STAGE 3 Standardized	Integration patterns are documented and reused. Provider selection moves into config with a written rationale per route — small-model-first with a fallback. Consumers call a published contract; a tiered router walks a fallback chain that gets logged.
STAGE 4 Integrated	The architecture assumes model-agnostic operation. A vendor change or a cheaper model is a config change, not a code change, and the seams hold when the transport layer moves underneath. Observability and budget limits are built in from the start.

PLACE YOUR ORG — circle where you are

- ○ **1 · Shadow** — AI is bolted in ad hoc, with providers hardcoded.
- ○ **2 · Sanctioned** — an AI integration pattern is written down and reused by more than one team.
- ○ **3 · Standardized** — provider selection lives in config with a fallback, behind a published contract.
- ○ **4 · Integrated** — swapping a model is a config change, and you can trace which provider served a call.

ASK YOURSELF

Could we change our model vendor next quarter without rebuilding?

AI can route across providers and log every hop. What cost, latency, and capability a problem actually needs is a human judgment.

THE MOVE

Build for the swap you know is coming.

Assume every provider and model will change, and build the seams that absorb it without a rewrite.

DIMENSION 10 OF 10

Team Capability & Culture

How AI fluency, learning, and psychological safety spread across a team.

STAGE 1 Shadow	AI fluency lives in individuals, not the team. A few people are good; most aren't, and there's no shared sense of what good looks like. One or two become the team's 'AI person.' The team can end up with output it can't explain — and when those people leave, the capability leaves too.
STAGE 2 Sanctioned	The org invests in basic AI literacy and some groups build informal habits — a real but uneven step. Underneath, status dynamics shift: junior engineers worry about looking dependent, senior engineers rework a craft they spent years building. It already shapes who feels safe asking for help.
STAGE 3 Standardized	AI literacy becomes part of onboarding, and teams write down their practices — when to use these tools, when to skip them, how to review output. The leader's real work here is building the safety to flag a confident-looking answer that's wrong without it costing status. Without that safety, the honest self-assessment everything else depends on never happens.
STAGE 4 Integrated	The team keeps adjusting its practices based on what works, and a new capability gets discussed before it spreads. Learning happens on the work itself — in review, in pairing, in arguing over a tricky output. The team also watches its own cost: reviewer load, the strain of supervising agents, the skills people worry about losing.

PLACE YOUR ORG — circle where you are

- ○ **1 · Shadow** — AI fluency sits with one or two people, not the team.
- ○ **2 · Sanctioned** — the org has put real money into AI literacy, however unevenly.
- ○ **3 · Standardized** — AI practices are written into onboarding, and it's safe to question AI output.
- ○ **4 · Integrated** — a new capability gets a team conversation before it spreads, and the team tracks its own load.

ASK YOURSELF

Is my team actually learning AI, or just leaning on one person who is?

AI can spread the mechanics, draft the practices, and do a lot of the first pass. Building the safety that lets someone ask the honest question out loud is human work.

THE MOVE

Make the honest question safe.

Build enough safety that someone can ask 'is this right?' out loud — then fluency becomes something the whole team can question and improve.

USING THE MODEL

Where most orgs are — and the first move from here

Across these ten dimensions, most organizations sit at Stage 1 or 2: AI is in wide use but only loosely permitted, and no one can quite say how well it's working. That isn't a failure — it's the normal early shape of adoption, where the tools arrived faster than the habits around them.

The distance to Standardized and Integrated is rarely about buying more tools. It's the unglamorous work: writing the standards down, putting review gates where the risk actually is, and keeping the people who own those standards close to where the work happens.

So the first move is small and specific: take the one or two dimensions you placed lowest, and move just those up a single stage. Each dimension's MOVE tells you how.

Come back to it in a few months.

Place your org again after a quarter or two. What's worth watching is which dimensions moved and which didn't — that gap, between what a team said it would do and what actually changed in the work, is usually where the honest conversation is.

Run this with your team.

Put the ten dimensions in front of the people who'd have to do the work. Place the org together, argue out the gaps, and leave with one or two dimensions you're actually going to move — and who owns each. That conversation, more than the score, is the point.

WORK WITH ME

Where I can help

I'm Ed Schaefer, an AI-Native Transformation Advisor. I've spent my career as a coach and a delivery leader, and I now help engineering organizations adopt AI in a way that holds up — governed, measured, and built into how the work actually gets done, without giving up the human judgment that has to stay with a person.

I work with leaders, executives, and teams on the parts of AI adoption that aren't really about the tools: the operating model, governance and guardrails, where human judgment belongs, and the culture and capability that make any of it stick. Most of that playbook already exists — in agile, lean, DevOps, and leadership development. AI just raises the stakes.

What that looks like in practice

Run this maturity assessment with your leadership team, and turn where you land into a plan.

Design the operating model and guardrails for AI-native engineering — the governance, the roles, the review gates.

Work through a single dimension that's stuck, with the leaders and teams who own it.

Get in touch

If any of this is live for your org, I'd be glad to talk — whether that's running the assessment together or working through the one dimension that's nagging you.

agile-ideation.com · linkedin.com/in/edward-schaefer