ENTERPRISE UX & PRODUCT STRATEGY LEADER

# The AI Adoption Playbook

A practical framework for governed, useful AI practice in the enterprise

Rusty Metty

rustymetty.com

## Executive summary

Most organizations don't have an AI problem. They have a judgment problem. Teams are picking up AI tools faster than they can evaluate what those tools produce — and the tools are very good at generating output that looks complete, looks confident, and looks done, whether or not it should be trusted.

The AI Adoption Playbook is a practical operating model for helping product, UX, and delivery teams use AI responsibly inside a large enterprise. It does not restrict experimentation. It gives every AI-assisted artifact a fixed structure to pass through before it reaches a real decision — a **harness**.

## The problem with scattered experimentation

Left ungoverned, AI adoption tends to follow the same pattern: individual designers and product managers pick up tools ad hoc, each develops their own private sense of what "good enough" looks like, and there is no shared standard for what an AI-generated artifact needs to prove before anyone acts on it.

The risk is not that AI produces bad output. The risk is that AI produces confident assumptions in the grammar of evidence — plausible, well-formatted answers that read as validated even when nothing has been validated at all. Without a shared harness, those assumptions reach roadmaps, executive decks, and customer-facing decisions unchecked.

## Harnesses: the core mechanism

A harness is a fixed evaluation structure wrapped around variable AI output. The idea borrows a sailing metaphor I use across this work: **standing rigging** is the fixed structure of a boat — the mast, the stays, the fittings that don't change from voyage to voyage. **Running rigging** is what varies — the sails you raise, the line you trim, the course you set for today's conditions. A harness applies the same split to AI-assisted work: the evaluation standard stays fixed. What flows through it changes every time.

**Running (variable) inputs** — what changes with every use:

- The artifact the AI produced

- The prompt and inputs fed to the system
- The context: product, team, and timing

**Standing (fixed) evaluation stages** — what never changes, run in sequence:

- **Load** — Name the artifact and the decision it's meant to serve. If no decision can be named, the evaluation stops here.
- **Reveal** — Document what the output revealed and what it obscured. Both columns are mandatory; an artifact that only reveals is being read uncritically.
- **Test** — Separate evidence from assumption. What's backed by real data, and what's a plausible-sounding guess wearing evidence's clothing?
- **Weigh** — Surface the consequences the artifact doesn't mention: human, business, and risk.
- **Judge** — Score against fixed criteria: evidence, human outcome, business fit, feasibility, trust and accessibility, differentiation. Scores don't average — a single "weak" on evidence or trust halts advancement regardless of how strong everything else scores.
- **Route** — Decide: Advance, Strengthen, or Stop. A logged Stop is a successful run of the harness, not a failed one.

## Prompting structures

Prompts are inputs to a system, not private improvisation. Structured prompting treats the prompt itself as something worth logging and reusing: what was asked, what context was given, and what constraints were set. This turns "I got a good result once" into a repeatable practice a team can build on.

## Evaluation criteria

Every harness run rests on the same discipline: evidence versus assumption. Evidence is something you could point to independently of the AI's output — a support ticket count, a finance-validated cost figure, a usage log. An assumption is something the artifact asserts that nobody has actually checked. Teams that skip this step tend to inherit the AI's confidence along with its guesses.

## Exploratory vs. production use

Not every AI-assisted artifact needs the same level of scrutiny. Sandboxed exploration — early ideation, internal brainstorming, throwaway prototypes — can move fast with a light harness. Anything moving toward a customer, a roadmap commitment, or a governance decision needs the full sequence, every time. The playbook's job is to make that line explicit before someone crosses it by accident.

## Human judgment checkpoints

At defined points in the process, a person signs off — not a model, and not a dashboard. The Judge and Route stages are deliberately human-owned. AI can help populate the Reveal and Test stages faster than a person could alone; it does not get a vote in what happens next.

## Governance across teams

The playbook works because it gives product, UX, engineering, security, and leadership a shared vocabulary for the same conversation. Instead of "is this AI thing okay to use," the question becomes "has this artifact passed Load, Reveal, Test, Weigh, and Judge — and what did Route decide?" That's a conversation every function can participate in on equal footing.

## Applying the harness: a worked example

An AI tool generates an opportunity canvas recommending self-service dispute resolution for a B2B billing platform. Run through the harness:

- **Load** — Decision: fund discovery next quarter, yes or no.
- **Reveal** — Revealed: dispute volume clusters in two account tiers; support cost per dispute runs 9x a self-service transaction. Obscured: assumes disputes are simple and rule-based; never mentions the collections team whose queue this feeds.
- **Test** — Evidence: support ticket volume and the cost multiple, both finance-validated. Assumption: "customers prefer self-service for disputes" — unvalidated.
- **Weigh** — Human consequence: high-emotion, money-in-dispute moments may be exactly where customers need a person, not a form. Business consequence: real savings if the complexity assumption holds, churn risk in the top-value tier if it doesn't.
- **Judge** — Business fit: strong. Human outcome: weak. Trust and accessibility: weak. One weak score on a fixed criterion is enough to hold the line.
- **Route** — Strengthen: run customer interviews on dispute complexity and escalation expectations before any roadmap commitment.

The business case was real. The harness is what kept it from shipping on an unvalidated assumption.

## Conclusion

AI adoption doesn't fail because teams move too fast. It fails when speed and judgment come apart — when an artifact reaches a decision before anyone has asked what it revealed, what it hid, and what it's asking you to assume. A harness doesn't slow teams down. It gives fast teams a fixed place to put their judgment, every single time. That's the throughline of this playbook, and of the broader work it sits inside: helping organizations turn ambiguity into executable systems people can trust.