The AI Governance Playbook

> **One-liner:** Design a practical AI governance framework for a team or project — covering when to use AI, how to verify outputs, and what requires human judgment.

? Jump in (Tinkerers start here)

Pick a real team, project, or organization you work with. You're going to design an AI usage framework they could actually adopt.

Step 1 — Map the AI touchpoints. Send this prompt:

> I'm designing an AI governance framework for a **[team type/project type]** that does **[describe the work]**. Map out all the places where team members might use AI in their workflow. For each touchpoint, classify the risk level:
>
> - **Low risk:** AI errors are easily caught and consequences are minor (e.g., drafting internal emails, brainstorming)
> - **Medium risk:** AI errors could waste significant time or create confusion (e.g., research summaries, data analysis, first drafts of client deliverables)
> - **High risk:** AI errors could cause reputational, legal, or financial harm (e.g., published content, financial recommendations, legal language, customer-facing decisions)
>
> Present this as a table with: Touchpoint | Description | Risk Level | Why

Step 2 — Design the verification tiers. Based on the risk map, create a tiered verification system:

> Based on the risk map above, design a 3-tier verification system:

> **Tier 1 (Low risk):** What's the minimum verification needed? What can proceed without review? **Tier 2 (Medium risk):** What checks are required? Who reviews? What's the turnaround expectation? **Tier 3 (High risk):** What's the full review process? Who signs off? What documentation is needed?
>
> For each tier, specify:
>
> - Verification steps (checklist)
> - Who is responsible
> - What "approved" looks like
> - What happens when issues are found

Step 3 — Write the team guidelines. Now produce the actual document:

> Write a 1-page "AI Usage Guidelines" document for this team. It should be practical, not corporate. Include:
>
> 1. **When to use AI** — Green light scenarios
> 2. **When to be careful** — Yellow light scenarios with required verification
> 3. **When NOT to use AI** — Red light scenarios or scenarios requiring explicit approval
> 4. **Verification standards** — The tier system from above, simplified
> 5. **Attribution** — When and how to disclose AI usage
> 6. **Escalation** — What to do when you're unsure whether AI use is appropriate
>
> Write it in the tone of a senior colleague giving practical advice, not a legal department issuing mandates.

Step 4 — Red-team the framework. Test it:

> Now role-play as a team member who wants to use AI in a gray area. Come up with 3 realistic scenarios where the guidelines are ambiguous or where a reasonable person might interpret them differently. For each scenario, suggest how to clarify the guideline.

Revise the guidelines based on the edge cases.

? Plan first (Planners start here)

Here's what you're about to do:

1. **Choose your context** — Pick a real team or project. The framework should be one you could actually share or implement.
2. **Map AI touchpoints and risk levels** — Identify every place AI could be used in the workflow and classify each by potential harm from errors.
3. **Design tiered verification** — Create different verification processes for different risk levels. Not everything needs the same scrutiny.
4. **Write the guidelines** — Produce a practical 1-page document that a team member could reference in their daily work.
5. **Red-team with edge cases** — Test the framework against ambiguous scenarios. Fix any gaps before sharing.

"Done" looks like: A complete, practical AI governance framework (risk map + tiered verification + 1-page guidelines) that you could present to your team, plus documentation of edge cases you tested against.

? Why this matters (Strategists start here)

Individual verification habits (from EP-Intermediate-01) don't scale to teams. When five people use AI differently with different standards, the team's AI output quality is only as good as the weakest link. A governance framework creates **shared standards without bureaucracy** — it tells people what's safe to do quickly and what requires care, without making every AI interaction feel like a compliance exercise. This is also the document organizations will pay for: a practical, calibrated AI usage policy that actually gets followed.

Reflection

- Which risk classification was hardest to assign? What does that ambiguity tell you?
- Would your team actually follow these guidelines? What would make them ignore it?
- Did the red-teaming step reveal fundamental gaps, or just edge cases?
- □□ *Present your framework to a colleague and ask: "Would you follow this?" Their honest reaction is more useful than any AI review.* (Social Learners)

?? Level up

You've reached the advanced level for Ethical Prompting & Judgment. From here, consider:

- Presenting this framework to your actual team and iterating based on feedback
- Combining this with AC-Advanced-01 to add governance to multi-agent workflows
- Building a case study of how the framework changed AI usage behavior in your team

Back to Ethical Prompting & Judgment

Revision #1
Created 2026-03-16 15:48:12 UTC by Admin
Updated 2026-03-16 15:48:12 UTC by Admin