1. What are we trying to learn?

What question are we exploring?

What uncertainty or assumption are we testing?

> ✎ *Example: Can AI-generated draft narratives reduce the time it takes to write audit observations?*

2. What's our small experiment?

What are we going to try?

How small can we make this? (1–2 weeks?)

> ✎ *Example: In the next sprint, we'll compare 3 audit findings written manually vs. with GPT-4 prompts in Word.*

3. Success looks like…

What outcomes will indicate this experiment is useful or worth scaling?

What metrics, signals, or artifacts will we look for?

> ✎ *Example: Observations are completed 25% faster and reviewers accept them with no additional editing.*

4. What might go wrong?

What risks, blockers, or ethical concerns do we need to anticipate?

What are our "safety rails"?

> ✎ *Example: Risk of hallucinated content; limit use to low-risk findings and use AI explainability prompts.*

5. What's our follow-up plan?

When and how will we evaluate the results?

Who needs to review the outcomes?

What's our next step if the results are positive (or negative)?

> ✎ *Example: Debrief at team sync on July 15; share findings with governance team for potential pilot scaling.*

📅 Timebox:

Start date: _____   End date: _____   Duration: ____ days