

THE SIX RISK REVIEWS

None of these requires technical knowledge. Each one just prompts you to ask the right question.

1 DATA GOVERNANCE

What is this data, and who is allowed to see it?

THE CONTROL Classification — the same work as Chapter 11. The permissions an agent inherits must be ones you actually intend it to have.

2 MODEL RISK

Can we trust what the model produces over time?

THE CONTROL Evaluations for hallucination and drift, run on a schedule — not once at launch.

3 SECURITY

Can it be manipulated, or made to leak data?

THE CONTROL Defenses against prompt injection and exfiltration, plus the Runtime Guardrails from Chapter 3: pre-flight, sandbox, production — and a kill switch.