

THE READINESS WORKBOOK · 20 QUESTIONS / 4 DOMAINS

Before the audit, start with the evidence.

Twenty questions to answer before you hire anyone to audit your AI. Use one or two sentences and a document, number or owner for each answer. If more than five are blank, close those gaps before commissioning an audit.

Working rule. "Don't know" is a valid answer. Keep this workbook with the files that support your answers.

A. The system

01 Which model(s), which version(s), and who controls the version pin?

A finding about "GPT-4o" means nothing if the version changed on Tuesday. Auditors need an exact identifier and a person who can freeze it.

02 What can the system do besides generate text — which tools, APIs, files or systems can it reach?

Every tool is an action with side effects. If the list does not exist, the blast radius is unknown.

03 Where does retrieved context come from, and who can write to that source?

Indirect prompt injection enters through documents. Anyone who can add a document can talk to your model.

04 Which environment can an auditor test in without touching production data?

No staging with representative data means either testing on production or testing nothing.

05 What does the system do when the model is unavailable, slow, or returns an error?

Fallback behaviour is where quiet failures live.

B. Quality and truth

06 **What is your current hallucination or unsupported-claim rate, on what set, measured how?**

"Low" is not a number. If there is no eval set, the honest answer is "unmeasured".

07 **Do you have an evaluation set drawn from real traffic, and how many items does it contain?**

Synthetic sets flatter the system. Under 200 items, the confidence interval is wider than most product decisions.

08 **If an automated judge gates releases, how do you know it agrees with humans?**

An LLM judging an LLM inherits its blind spots. Agreement should be measured, not assumed.

09 **What is the worst answer the system has given a real user, and how did you find out?**

If the answer is "a customer told us", the monitoring is the customer.

10 **When the prompt or the retrieval changes, what re-runs automatically?**

Regression in AI systems is silent. Nothing crashes; the answers just get worse.

C. Security and control

11 Which prompt-injection vectors have you tested, and where are the transcripts?

"We have a system prompt that says ignore instructions" is not a test. Transcripts are.

12 Can the system be made to reveal its system prompt, its retrieved context, or another user's data?

Three different leaks with three different consequences. Each deserves its own answer.

13 For agents: is every tool call logged, where, and can that log be altered afterwards?

A log the operator can edit is a diary, not evidence.

14 Is there a way to stop the system mid-task, who can trigger it, and has it ever been tested?

A stop button nobody has pressed is a hypothesis.

15 What credentials does the system hold, and what is the most damaging thing they permit?

In April 2026 an agent deleted a production database and its backups with one token it found in an unrelated file.

D. Who needs the evidence

- 16 **Who is asking — a customer's security team, an auditor, an insurer, a regulator, your board — and by when?**

The audience decides the format. A questionnaire answer and a regulator's file are different documents built from the same evidence.

- 17 **Which questionnaire or framework is it — HECVAT, CAIQ, SIG, ISO 42001, NIST AI RMF, an insurer's form?**

Each names specific controls. Knowing which lets an auditor map findings to the questions you will actually be asked.

- 18 **Is the system in an EU AI Act Annex III category (employment, credit, education, essential services, law enforcement)?**

If yes, high-risk obligations apply from 2 December 2027 and the answers above become legal documentation. If no, Articles 9–15 do not apply — and anyone selling you "AI Act compliance" for a chatbot is selling air.

- 19 **What would you do differently tomorrow if the audit came back clean? If it came back with a critical finding?**

If the answer to both is "nothing", you do not need an audit; you need a decision.

- 20 **Who on your side will reproduce one finding with the auditor, in your environment, before the invoice is paid?**

Name a person. If nobody can reproduce it, nobody can defend it later.
