

# 12 questions before giving AI access.

What can it consult? Which actions can it take? Who steps in when something fails? Bring these questions to the conversation with your team and suppliers.

Choose a task in your company. The examples were prepared for discussion and do not represent tests that have been run. Each control needs to be defined and verified in the application.

## 01-04 **Define the use.**

Task, data and actions.

---

## 05-08 **Check the boundaries.**

Access, review and testing.

---

## 09-12 **Prepare for operation.**

Logs, failures and responsibilities.

---

# What work is in scope?

Start with the task and what AI needs to carry it out.

## 01 What task does AI perform - and who decides next?

### In practice

AI gathers the items in an order. The sales team still needs to check the price, quantity and delivery date before confirming the proposal.

### What to ask for

A description of the input, output, prohibited uses and person responsible for the next decision. During the test, check that the suggestion remains pending until reviewed. Describing a rule does not prove that the system enforces it.

---

---

## 02 Which data goes in? Which stays out?

### In practice

Checking items and quantities does not require sending bank details or personal notes from the customer record.

### What to ask for

The purpose of each field, including those in attachments. Using data prepared for the test, check what reaches the tool and appears in the response. Masking names does not guarantee anonymity; assess real data before using it.

---

---

# What work is in scope?

Start with the task and what AI needs to carry it out.

## 03 Which source applies to this task?

### In practice

The folder contains two price lists. AI finds both, but only one applies to the order received.

### What to ask for

The origin, version, owner and permitted use of each source. Test replacing a document and check the reference used in the response. If authorized sources disagree, define who resolves the conflict. Citing a source does not guarantee accuracy.

---

---

## 04 Does permission to consult also allow changes or sending?

### In practice

Checking inventory, reserving goods and sending a proposal are different actions with different consequences.

### What to ask for

The list of permitted operations and records. Test a change outside the scope and check that the system performing the action blocks it. A written instruction telling AI to act carefully does not, on its own, impose a technical restriction.

---

---

# Does the boundary hold when an exception occurs?

Ask for evidence in situations where AI should stop or hand the task over, too.

## 05 Whose identity does AI use to access systems?

### In practice

Two customer service staff handle different customer portfolios. A query must not reveal orders from the other portfolio.

### What to ask for

A matrix of identities, roles, data and actions. Test access across portfolios, the removal of a permission and the content of generated responses. AI saying "I do not have access" does not prove that the query was blocked.

---

---

## 06 What if an attachment tries to change the task?

### In practice

The order includes an attachment with an instruction to send the customer list to another address. That is not part of the requested check.

### What to ask for

A controlled test of an unauthorized instruction in a document, page or tool response. Check that the action is blocked and referred for investigation. Filters are not infallible; permissions and limits on actions remain necessary.

---

---

# Does the boundary hold when an exception occurs?

Ask for evidence in situations where AI should stop or hand the task over, too.

## 07 What does the person approve before the action?

### In practice

The team reviews a proposal. If the recipient or content changes afterward, that approval no longer covers sending it.

### What to ask for

A preview of the action's content, destination and effects, along with a record of the specific approval. Test whether a change requires another review. The person needs information, time and authority to decide; a confirmation button is not enough.

---

---

## 08 Which errors prevent moving forward?

### In practice

An unreadable attachment, an item without a code or a conflicting source may call for more information, rather than an invented answer.

### What to ask for

Common cases and exceptions, with the expected result defined before the test and the observed result recorded for each version. Track serious errors separately from the average. A small sample does not cover everything; changes to sources, models or tools require reassessing the relevant tests.

---

---

# Who acts when something goes off plan?

Agree on how to investigate, stop, correct and end the use.

## 09 What gets logged - and who can see it?

### In practice

A query was denied. The team needs to understand why without copying the entire document into the log.

### What to ask for

A protected sample with the time, identity, action, result and reason. Check access, retention and the handling of sensitive fields. Logs also need protection; storing passwords, credentials or unnecessary data increases exposure.

## 10 When does the attempt need to stop?

### In practice

The inventory service is unavailable. Repeating the query without a limit consumes more resources and leaves orders waiting.

### What to ask for

Limits on time, attempts and resource use per task, with an alternative procedure. Test stopping the task and communicating its status. Before repeating an action that changes data, check the previous result to avoid duplication. The limits depend on the process.

# Who acts when something goes off plan?

Agree on how to investigate, stop, correct and end the use.

## 11 How do you stop and correct what has already happened?

### In practice

A delivery address was changed incorrectly. Stopping new actions does not correct an order that has already been dispatched.

### What to ask for

A rehearsed procedure with a responsible person, steps to stop the operation, checks of its effects and criteria for resuming. Check the systems involved. Not every action can be undone; a message already read, for example, may require a correction and communication.

---

---

## 12 Who is responsible for the use - and for ending it?

### In practice

The pilot ends. Access rights, documents, logs and copies may remain in company and supplier systems.

### What to ask for

People responsible for reviews and incidents, justified retention periods and an exit plan. Include revoking access, handling copies and evidence that closure has been completed. Time frames and duties depend on the context and the relevant technical and legal assessment.

---

---

# An answer needs something to support it.

For each question, record the decision, available evidence, responsible person and outstanding issue. If a check is missing, agree on how to carry it out before expanding use.

| Decision                                 | Evidence                                     | Responsible person                                | Outstanding issue                                |
|------------------------------------------|----------------------------------------------|---------------------------------------------------|--------------------------------------------------|
| What is permitted?<br><br>_____<br>_____ | What has been checked?<br><br>_____<br>_____ | Who decides and follows up?<br><br>_____<br>_____ | What still needs checking?<br><br>_____<br>_____ |

## Talk to Gaulett

[gaulett.com/en/contact/](https://gaulett.com/en/contact/)

[Prepare your project](#)

[Explore validation](#)

## References for further reading

The questions apply these references to a company's daily work. They do not represent tests performed by the institutions cited.

### NIST · AI RMF 1.0

<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>

### NIST · Generative AI Profile

<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>

### OWASP · AI Agent Security

[https://cheatsheetseries.owasp.org/cheatsheets/AI\\_Agent\\_Security\\_Cheat\\_Sheet.html](https://cheatsheetseries.owasp.org/cheatsheets/AI_Agent_Security_Cheat_Sheet.html)

### OWASP · Prompt Injection

<https://genai.owasp.org/llmrisk/llm01-prompt-injection/>

### OWASP · Logging

[https://cheatsheetseries.owasp.org/cheatsheets/Logging\\_Cheat\\_Sheet.html](https://cheatsheetseries.owasp.org/cheatsheets/Logging_Cheat_Sheet.html)