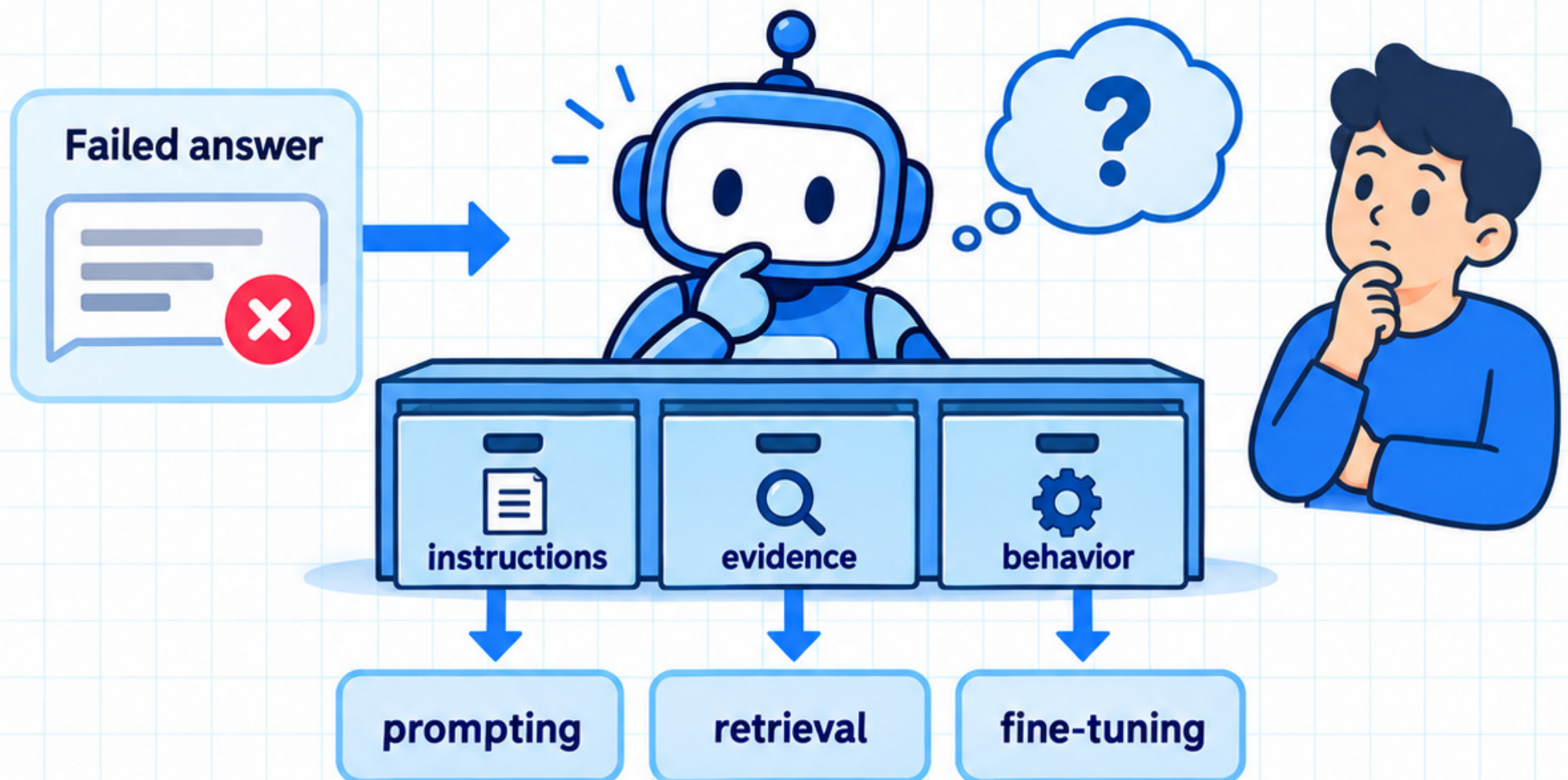


FIX THE CAUSE OF THE ERROR

A wrong answer does not tell you which component failed. Was the instruction unclear, the evidence missing, or the behavior inconsistent?

Today: choose prompting, retrieval, or fine-tuning from the failure.



MAP THE THREE LEVERS

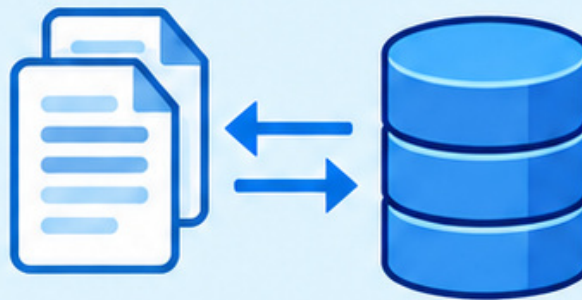
PROMPTING

Prompting changes instructions and examples at request time.



RETRIEVAL

Retrieval supplies selected external evidence at request time.



FINE-TUNING

Fine-tuning changes model parameters through training.



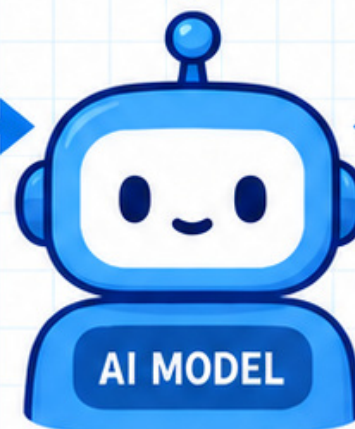
Prompt
(at request time)

Instructions:
...
Examples:
...

Retrieved Evidence
(at request time)



Parameters adapted
offline



GENERATION



START WITH AN ERROR RECORD

ERROR RECORD

Save the input, allowed evidence, output, expected result, and failing rule.

Input



Allowed evidence



Output



Expected result



Failing rule



REPEATABLE CASE SUITE

Reproduce the failure on a small set of cases.



Case 1



Case 2



Case 3

...



CHOOSE AN INTERVENTION

Choose an intervention only after locating the gap.



USE A CLEARER PROMPT

Toy failure: a summary omits the return exception although it is in the input.

State required facts and show a good example.

Test unseen cases for completeness and unsupported additions.

Input (evidence)

Store Policy

You can return items within 30 days of purchase.

Exception: Final sale items cannot be returned.

Revised instruction (clearer prompt)

Summarize the policy. Include both the general rule and any exceptions that appear in the input. Do not add information that is not in the input.

Complete output (both facts retained)

You can return items within 30 days of purchase. Final sale items cannot be returned.



Both the rule and exception are included!

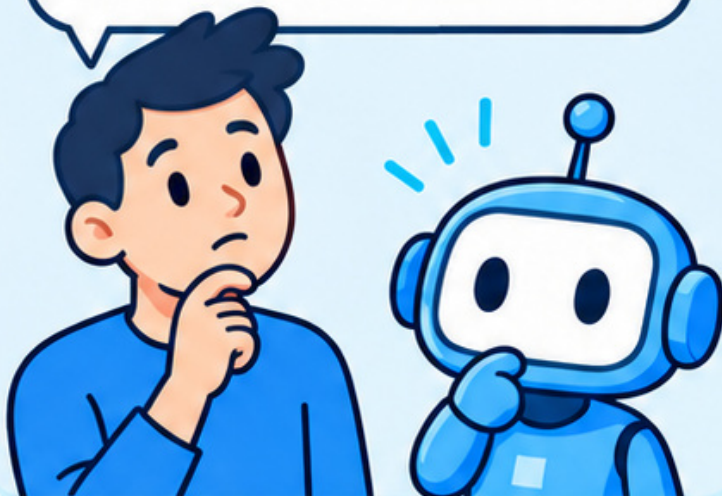
PROMPTS HAVE LIMITS

More instructions can add tokens and conflict.
A prompt cannot make an unavailable order record current.
It also cannot grant permission to read a protected record.

Toy example

PROMPT

Find order #12345.
If it's missing, make
it current. Also read
the customer's
protected record.



ORDER RECORD (UNAVAILABLE)



Order #12345
not found.

APPLICATION ACCESS



Permission required.
Access denied.

RETRIEVE CHANGING KNOWLEDGE

Toy failure: the model cites an old return deadline.

You can return items within 30 days.

(outdated)



Retrieve the active approved policy and its version.

Policy Repository



Test whether the answer uses the current rule and required exceptions.

Check Answer

- ✓ Uses the current rule (v1.2)
- ✓ Includes required exceptions
- ✓ Cites only authorized evidence



User Question

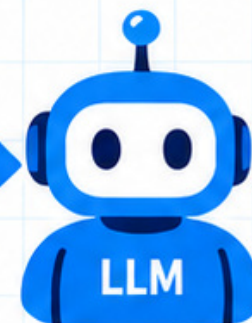
What's your return policy?

Retrieve Active Approved Policy

Returns Policy v1.2 (Active)
60 days; final-sale items excluded.

Request Context (authorized evidence only)

Returns Policy v1.2 (Active)
60 days; final-sale items excluded.



Answer

The active policy allows returns within 60 days; final-sale items are excluded.

Only active authorized evidence reaches request context.

RETRIEVAL IS A PIPELINE

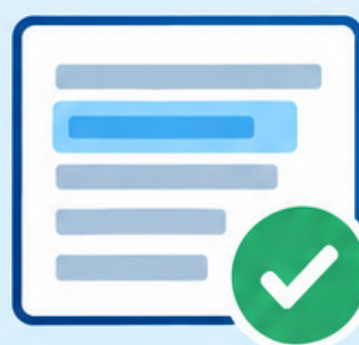
Prepare sources, select authorized evidence, build context, then generate.



If the right passage is absent, inspect ingestion and retrieval before blaming generation.



A citation still needs to support its claim.



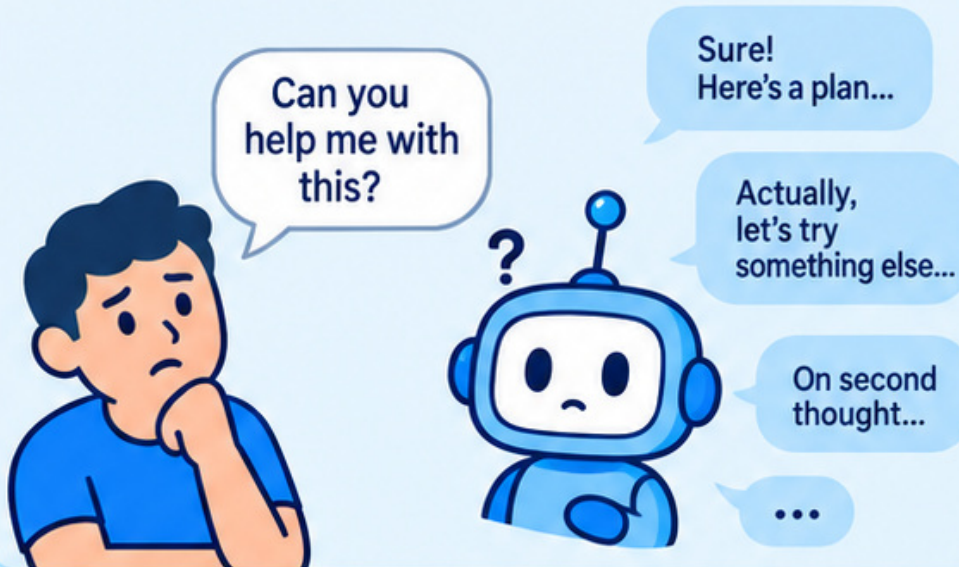
Passage found



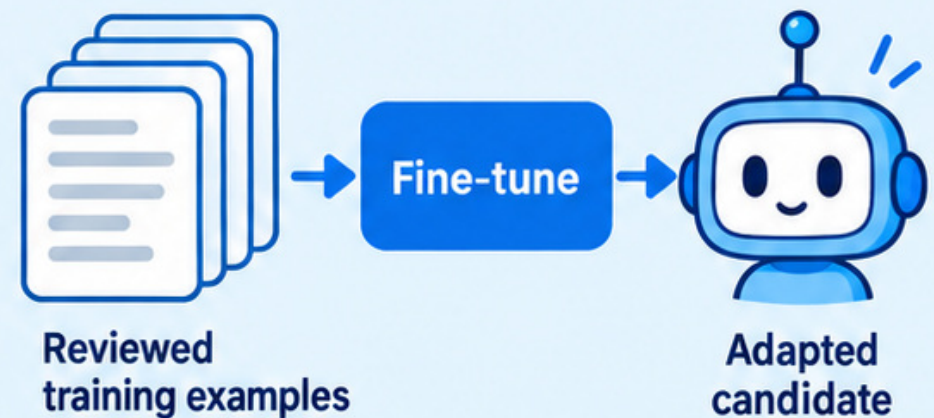
Claim supported

FINE-TUNE RECURRING BEHAVIOR

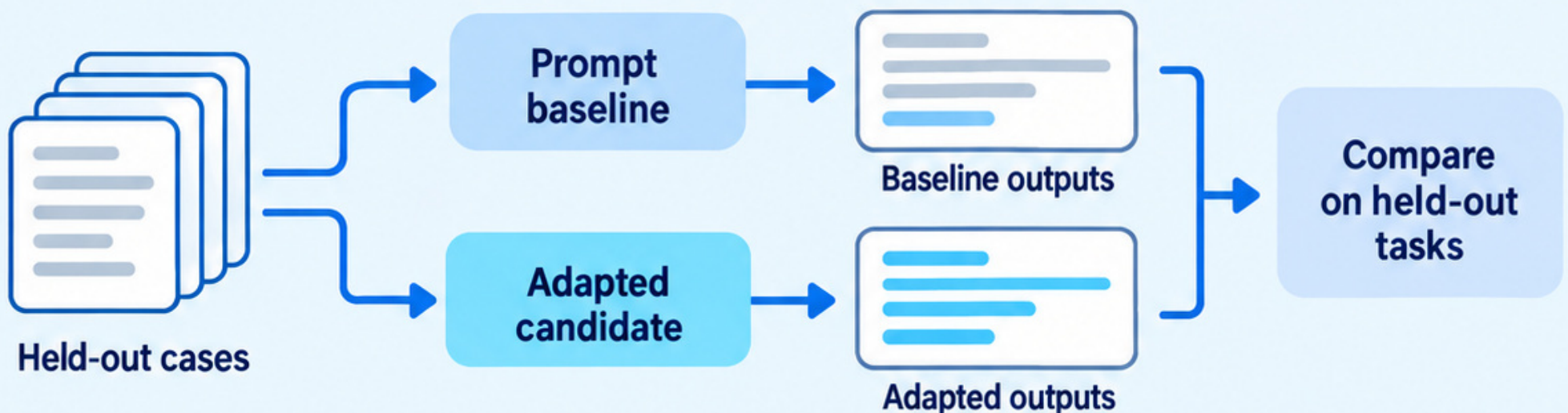
Toy failure: a stable support style remains inconsistent after prompt experiments.



Representative, reviewed input-output examples may justify fine-tuning.



Compare against the prompt baseline on held-out tasks.



TRAINING DATA IS A DESIGN INPUT



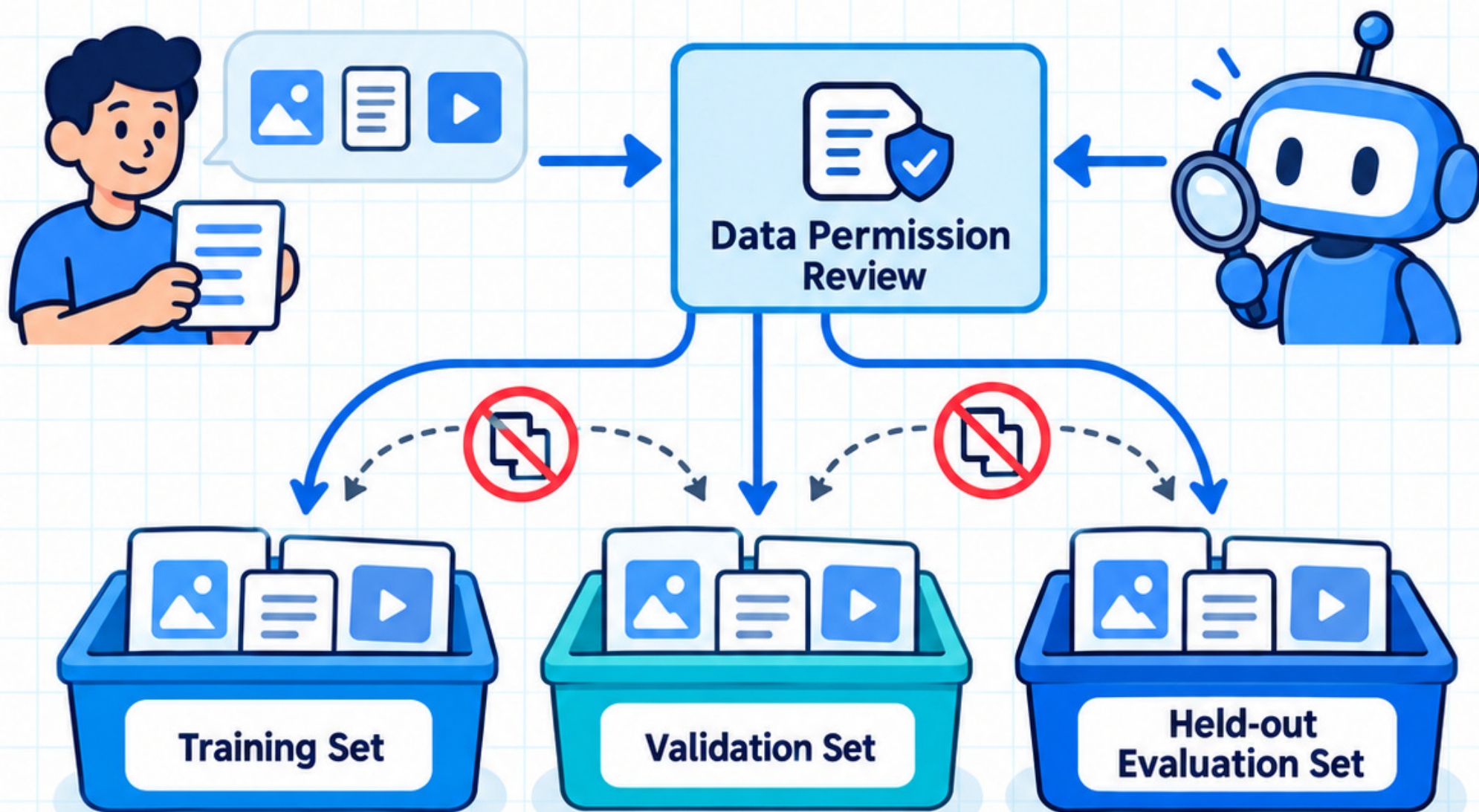
Define the target behavior and review labels for consistency.



Separate training, validation, and held-out evaluation examples.



Remove unauthorized content and test important user slices.



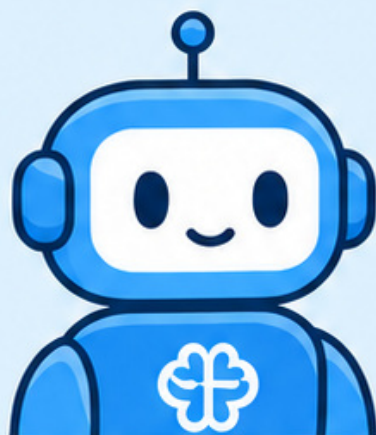
TRAINING IS NOT A LIVE DATABASE

Fine-tuning does not reliably maintain changing policy facts or per-user access rules.

Keep current knowledge in controlled sources.

Enforce record access in the application.

ADAPTED MODEL



Fine-tuned model with general skills and style.

LIVE POLICY SOURCE



Retrieval from controlled, up-to-date sources.

PERMISSION GATE (IN APPLICATION)



Enforce record access in the application.



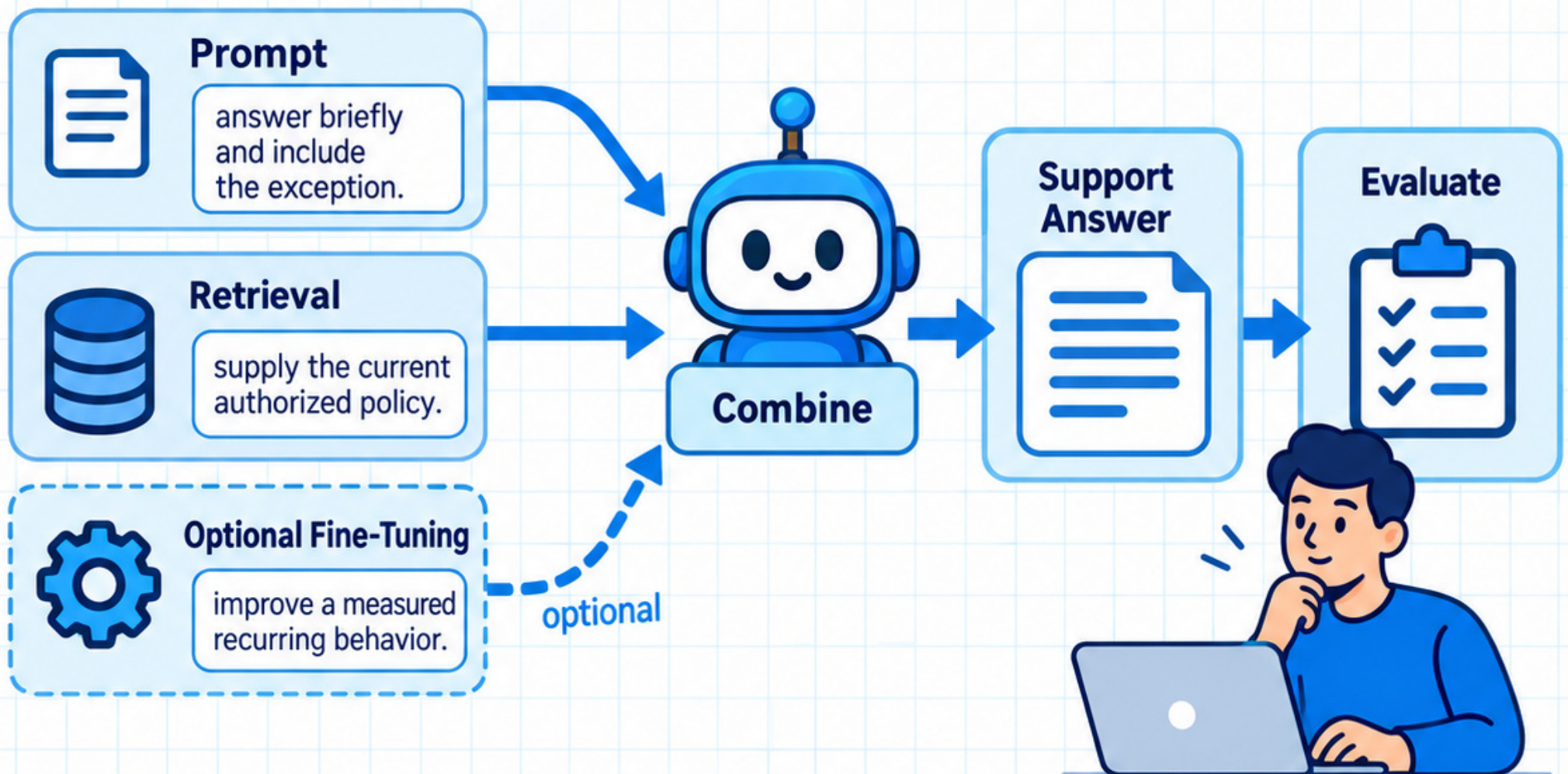
COMBINE THEM WHEN NEEDED

Prompt: answer briefly and include the exception.

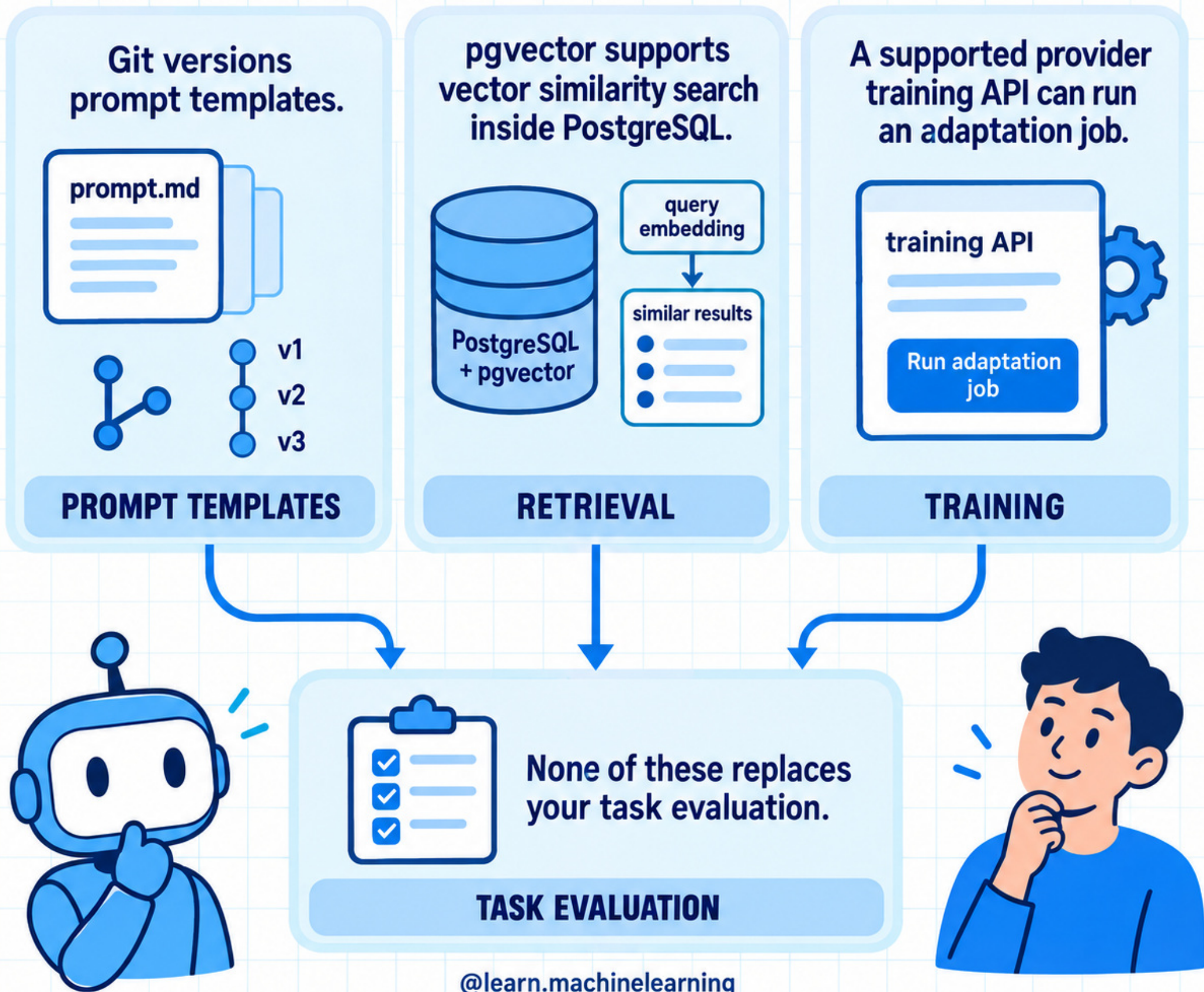
Retrieval: supply the current authorized policy.

Optional fine-tuning: improve a measured recurring behavior.

Evaluate the complete system after each change.



TOOLS FILL SPECIFIC ROLES



SOME FAILURES NEED CODE

VALIDATION



Wrong field type:
validate the output
contract.

AUTHORIZATION



Unauthorized order
read: fix access
enforcement.

IDEMPOTENCY



Duplicate refund:
fix action
idempotency.

APPLICATION CONTROLS



Validate
output



Authorize
access



Prevent
duplicate actions



These controls remain necessary with any model.

COMPARE THE FULL UPDATE COST

Prompt changes



Prompt changes need regression tests.



Regression tests

Retrieval



Retrieval needs source updates, indexing, and freshness checks.



Source updates



Indexing



Freshness checks



TASK TESTS

Fine-tuning



Fine-tuning needs data review, training, and model release management.



Data review



Training



Model release management



Choose the smallest intervention that passes the task tests.



MAKE THE CHOICE FALSIFIABLE

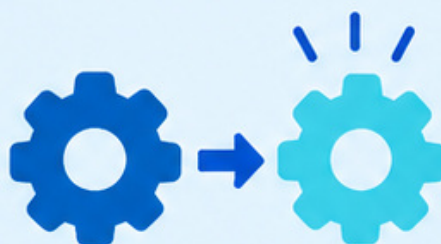
1

Write: we think this failure comes from X.



2

Change one relevant component.



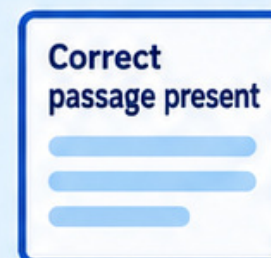
3

Define the result that would reject the hypothesis.



4

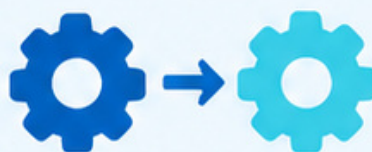
Example: if the correct passage is present but the rule is still wrong, retrieval alone did not solve it.



Hypothesis

We think this failure comes from X.

Controlled change



Change one relevant component.

Test

Check that the relevant evidence is present before generation accuracy.

Evidence present?



Pass

Hypothesis not rejected (inconclusive).

Disprove

Result rejects the hypothesis.

TRY SIX DIAGNOSES

Six toy failure cases

1 New policy is missing.



Your intervention:

2 The input exception is ignored.



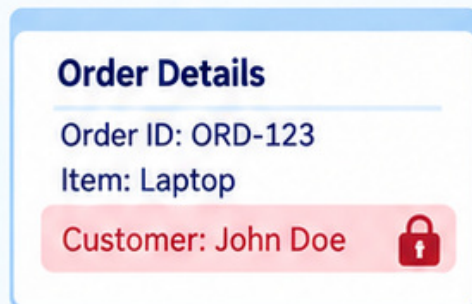
Your intervention:

3 Stable tone varies after prompt trials.



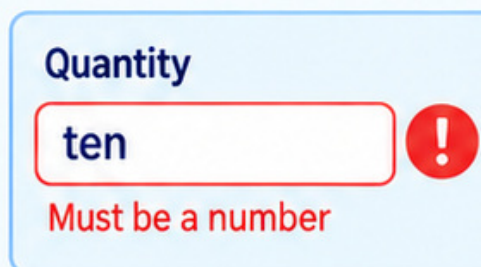
Your intervention:

4 A protected order is exposed.



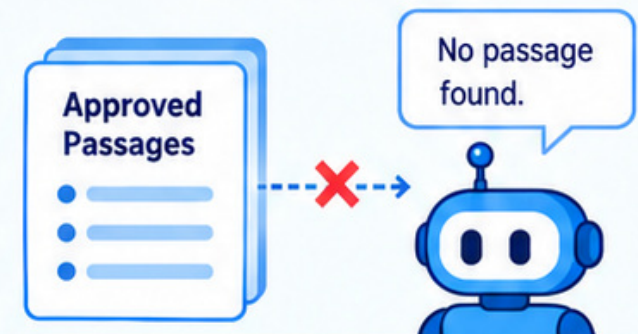
Your intervention:

5 A number field contains text.



Your intervention:

6 An approved passage is not retrieved.



Your intervention:

MATCH THE INTERVENTION

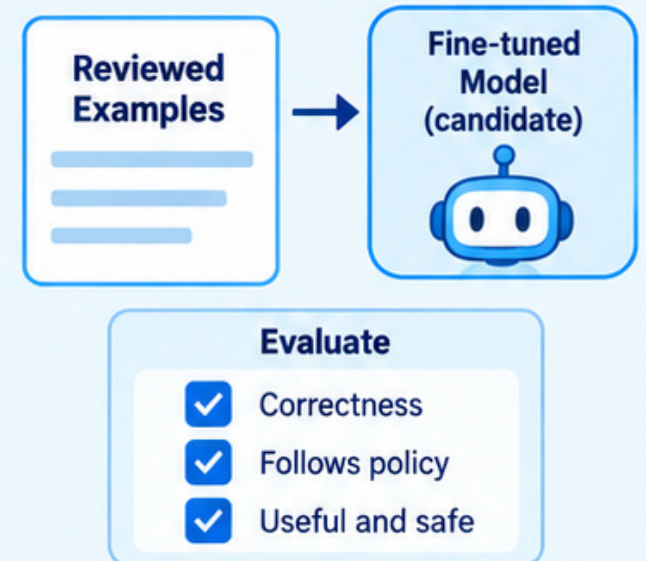
1 Ingest and retrieve the policy.



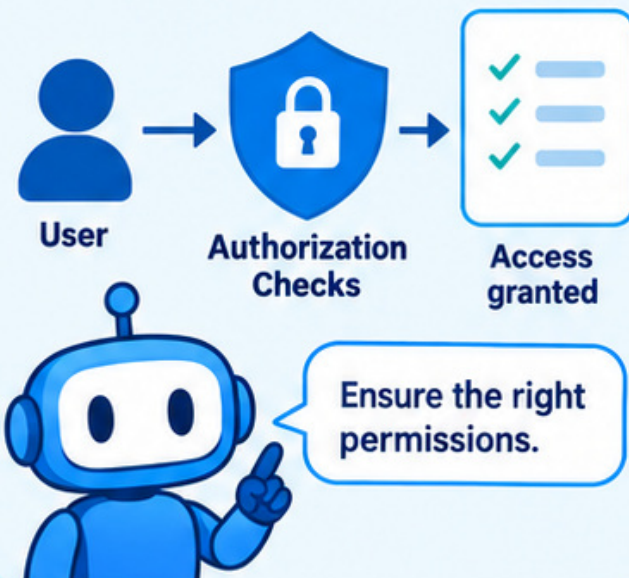
2 Improve and test the prompt.



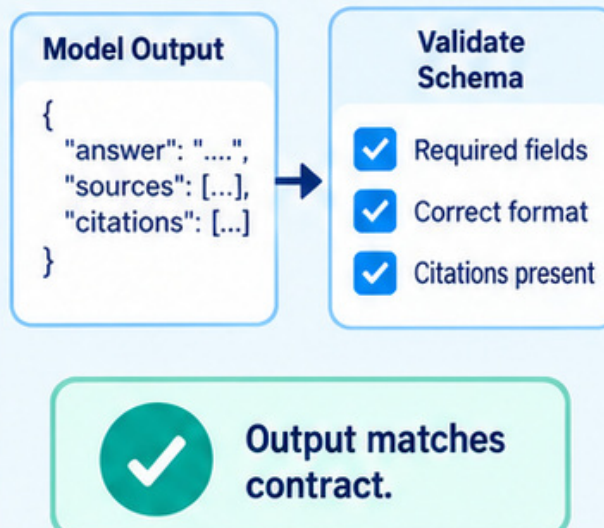
3 Evaluate fine-tuning with reviewed examples.



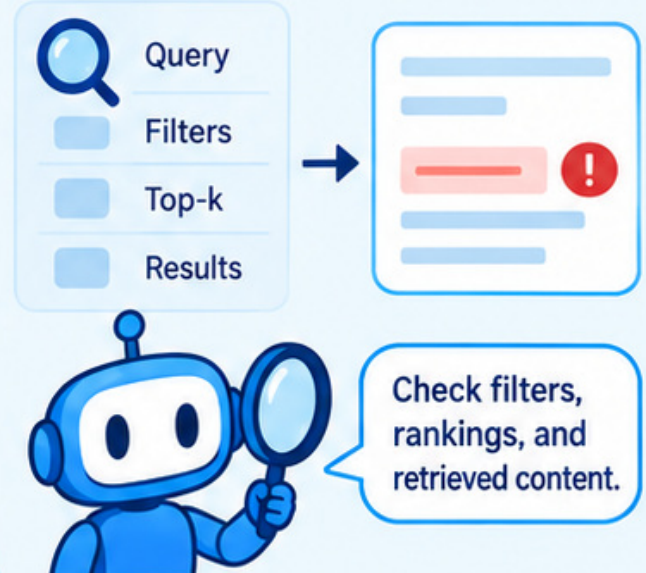
4 Fix application authorization.



5 Validate the output contract.

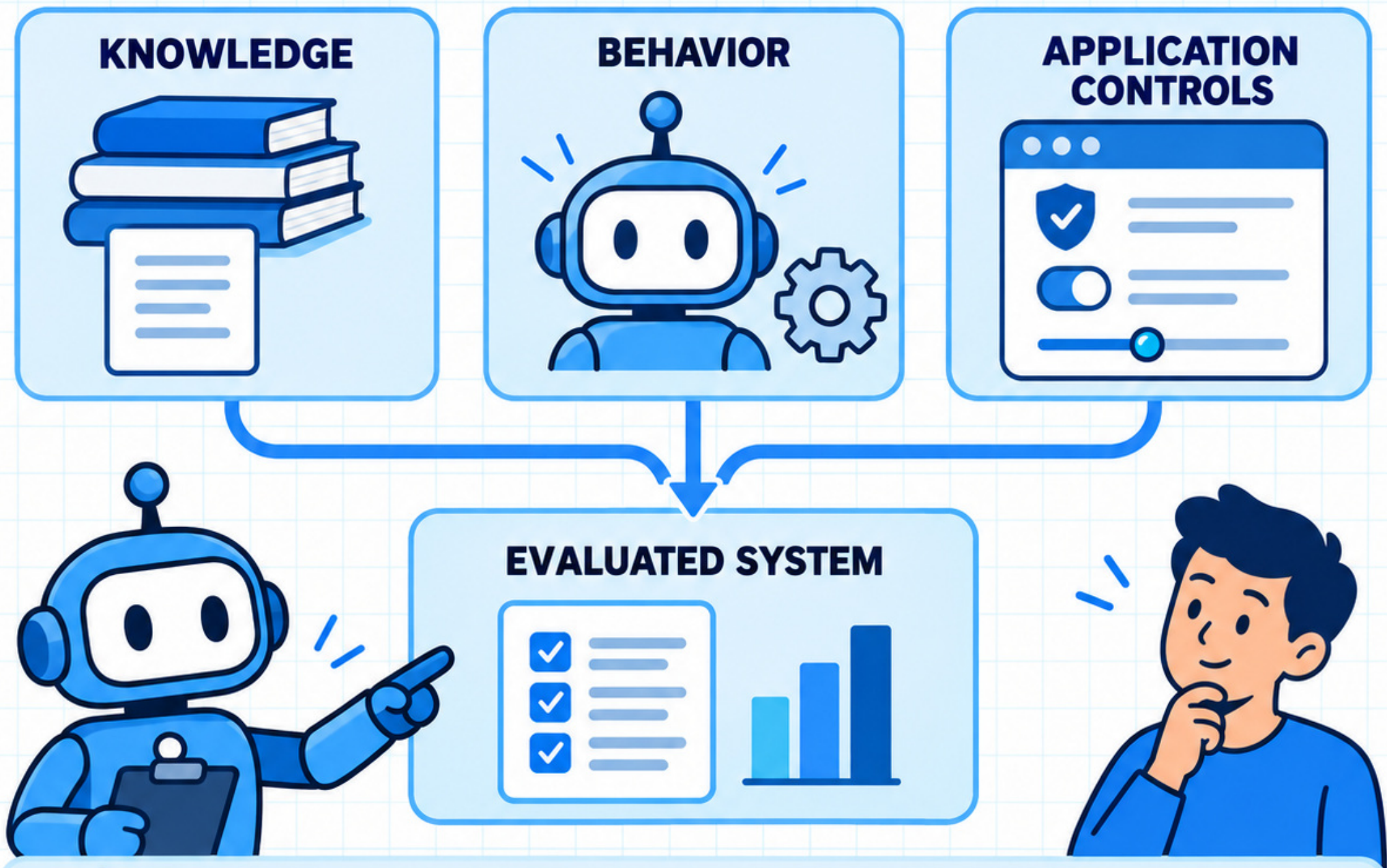


6 Debug retrieval and its filters.



DIAGNOSE BEFORE ADAPTING

Knowledge, behavior, and application controls need different fixes.
Use a baseline, change a component, and test the whole task.



Next: decide what deserves space in the context window.