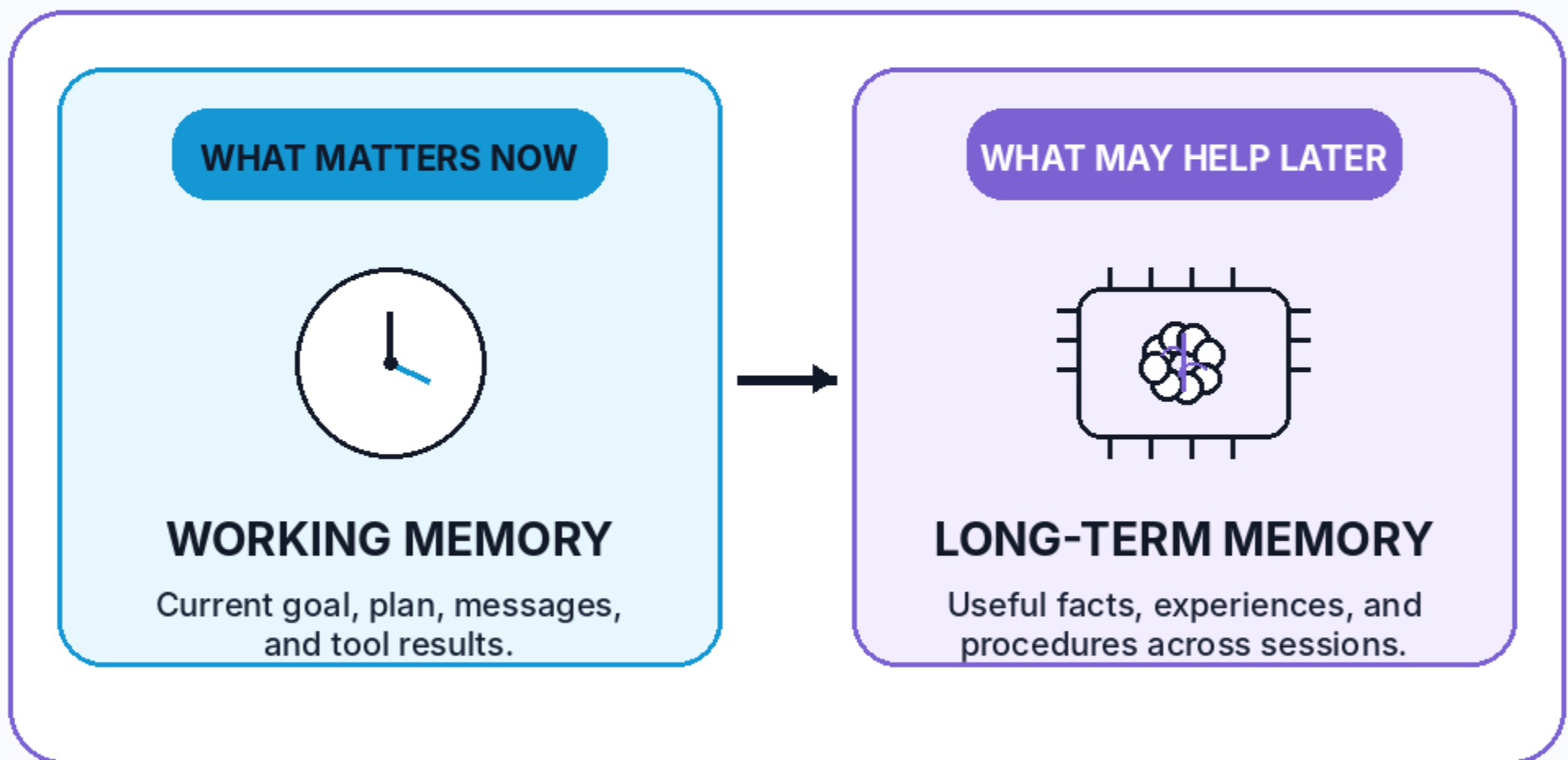


HOW AI AGENTS REMEMBER

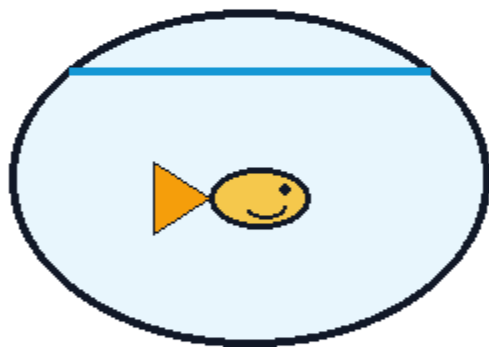
Memory is more than chat history—and more is not always better.



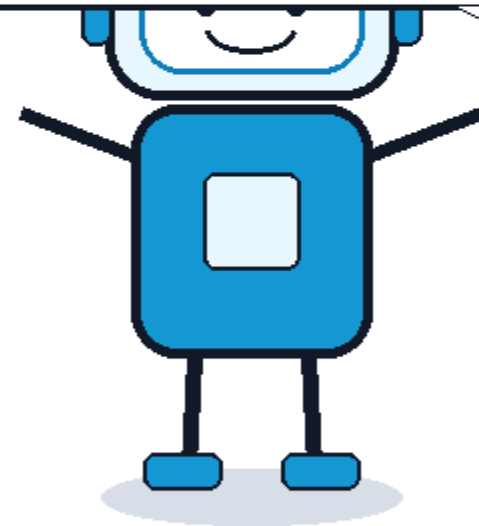
Good memory is selective, relevant, correctable, and safe.

MEET THE GOLDFISH AGENT

It can be smart in the moment—but forget everything as soon as the session ends.



"You already told me your preference... but I cannot remember it now."



✗ Repeats questions

"What is your timezone?"—again.

✗ Loses progress

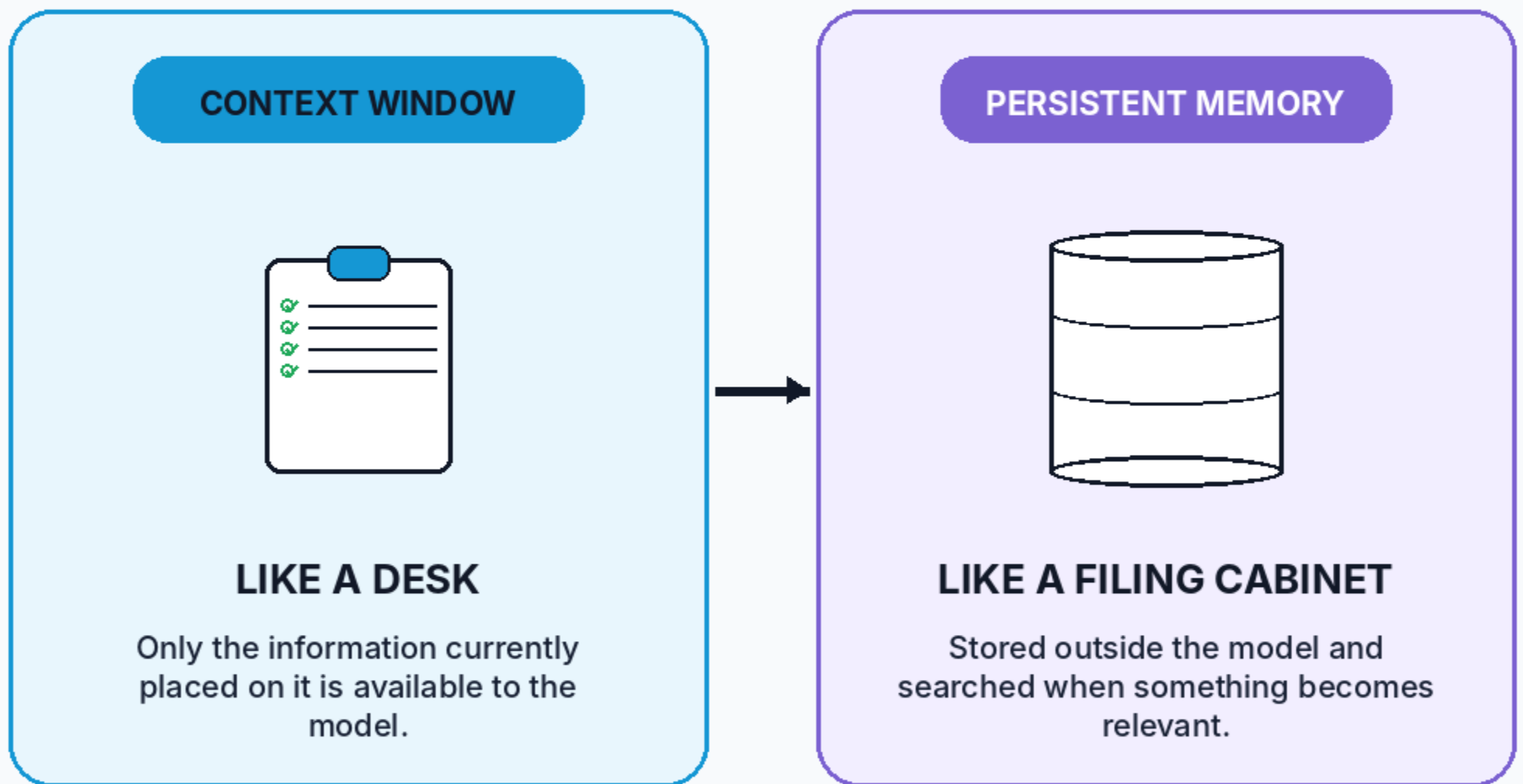
Forgets which steps are already complete.

✗ Cannot personalize

Treats every interaction like the first one.

MEMORY IS NOT THE CONTEXT WINDOW

Context is what the model can see right now. Persistent memory lives outside and must be retrieved.



Memory helps only when the right item is brought back into context.

TWO MEMORY TIME HORIZONS

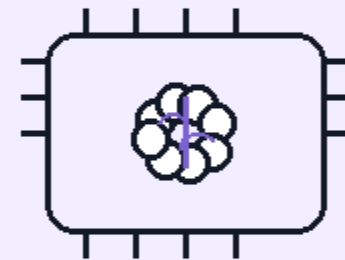
Agents usually need both: continuity within a task and continuity across sessions.

SHORT-TERM / WORKING



- ✓ Current conversation
- ✓ Current plan and progress
- ✓ Recent tool results
- ✓ Temporary notes for this task

LONG-TERM



- ✓ Useful user preferences
- ✓ Past events and outcomes
- ✓ Rules, procedures, and skills
- ✓ Information across sessions

Short-term remembers the task. Long-term remembers what may help future tasks.

WORKING MEMORY: THE AGENT'S SCRATCHPAD

It keeps the agent oriented while a task is still in progress.

GOAL: PLAN A 3-DAY TRIP

1



USER REQUEST

Jaipur • ₹30,000 • vegetarian food

2



CURRENT PLAN

Flights → hotel → itinerary → budget check

3



LATEST RESULT

Morning flight is sold out

4



NEXT STEP

Search afternoon flights

When the task ends, most scratchpad details can be summarized—or discarded.

THREE USEFUL TYPES OF LONG-TERM MEMORY

A simple human analogy helps: facts, experiences, and rules.

SEMANTIC



FACTS & PREFERENCES

What is true or useful about the user or world?

EPISODIC



PAST EXPERIENCES

What happened, in what order, and with what result?

PROCEDURAL



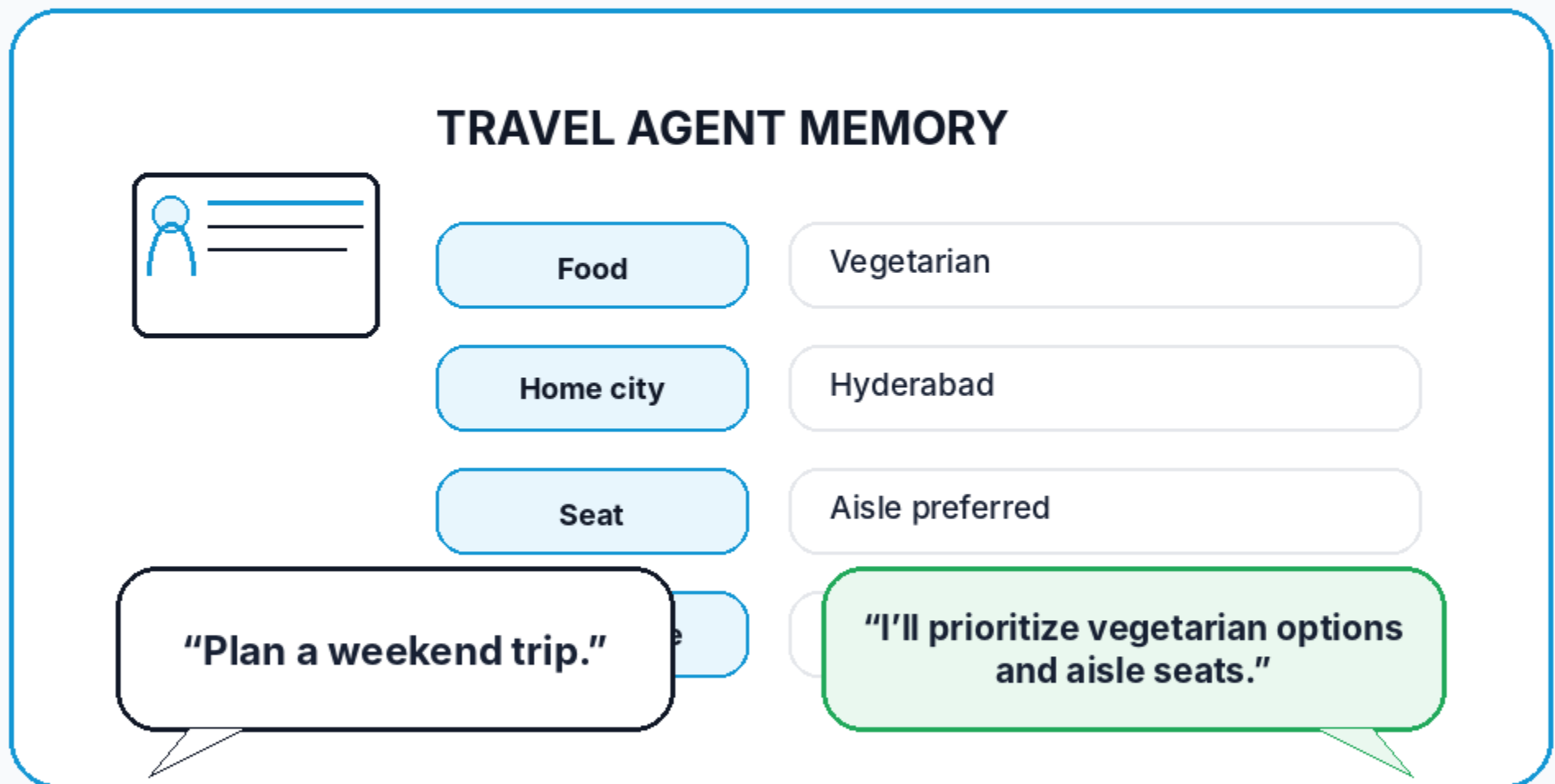
RULES & SKILLS

How should this kind of task be performed?

An agent may use one type—or combine all three.

SEMANTIC MEMORY: FACTS THAT STAY USEFUL

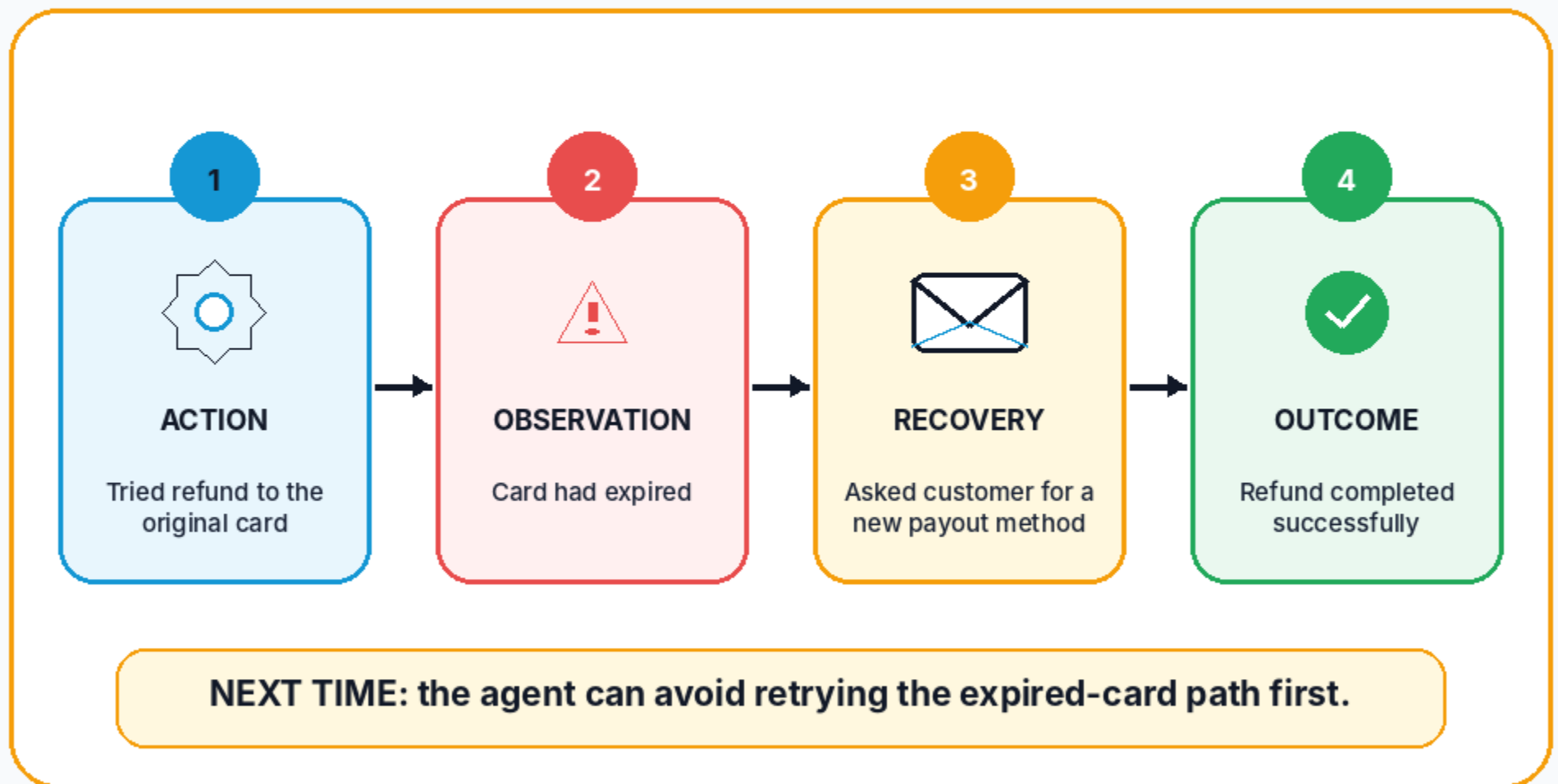
This is the “what I know about you or the task” memory.



Good semantic memory reduces repeated questions and supports personalization.

EPISODIC MEMORY: WHAT HAPPENED BEFORE

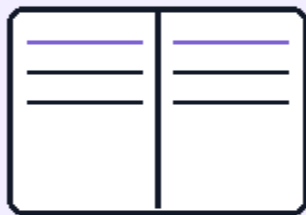
It remembers an experience—not just the final fact.



Episodic memory can preserve successful—or failed—sequences and their outcomes.

PROCEDURAL MEMORY: HOW THE WORK SHOULD BE DONE

Think of it as an onboarding guide, checklist, or reusable skill.



COMPANY REFUND PROCEDURE

- ✓ Verify order ownership
- ✓ Check policy eligibility
- ✓ Ask approval above ₹5,000

"I know the approved steps and where human approval is required."

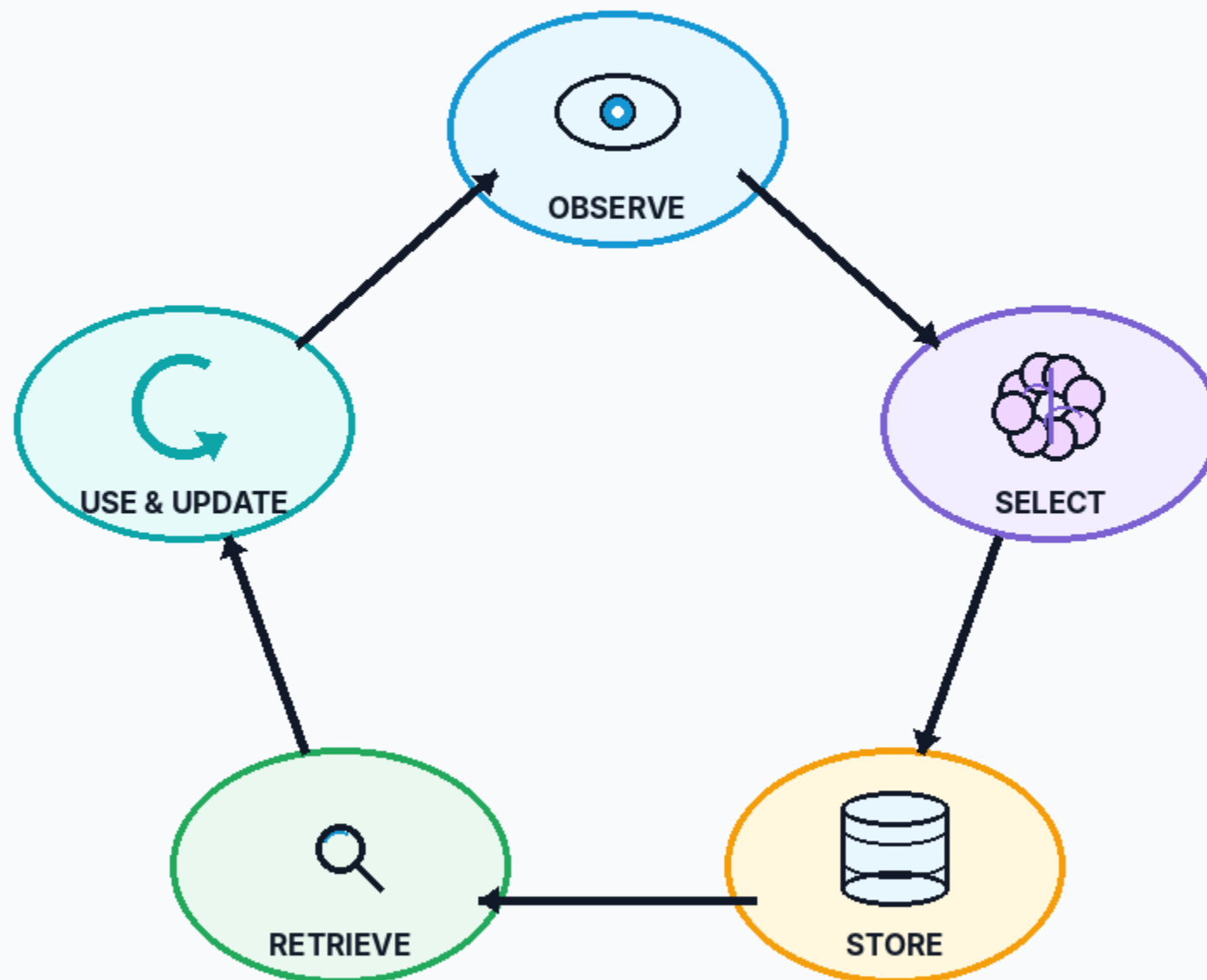


REUSABLE SKILL

Procedures should be versioned and reviewed—not silently treated as permanent truth.

THE MEMORY LIFECYCLE

An agent should not save every sentence. It needs a controlled process.



FORGET OR DELETE when memory is stale, sensitive, incorrect, or no longer needed.

SHOULD THE AGENT REMEMBER THIS?

Use a simple three-question filter before writing long-term memory.

1



WILL IT HELP LATER?

A stable preference or reusable lesson may be useful.

2



IS IT RELIABLE?

Prefer confirmed facts over guesses or one-off statements.

3



IS IT ALLOWED?

Respect consent, privacy, retention rules, and user control.

If the answer is "no" or "not sure," keep it temporary—or do not store it.

MEMORY NEEDS A GOOD LIBRARIAN

Dumping every past detail into the prompt can make an agent slower, costlier, and more confused.

BAD RETRIEVAL



"LOAD EVERYTHING"

Relevant facts get buried inside noise and stale information.

GOOD RETRIEVAL

"Bring only the memories relevant to this user and this task."



SEARCH → FILTER → RANK →
USE

The goal is not maximum memory. It is the right memory at the right time.

FULL EXAMPLE: A SUPPORT AGENT WITH MEMORY

One customer, one issue—four kinds of useful state.

GOAL: RESOLVE A REPEATED DELIVERY ISSUE

WORKING



Current order #8421 is delayed by 3 days.

SEMANTIC



Customer prefers replacement over store credit.

EPISODIC



Last replacement went to the old address.

PROCEDURAL



Address changes require confirmation before shipment.

RESULT: confirm the new address first, then create the replacement.

MEMORY VS RAG: RELATED, NOT IDENTICAL

Both retrieve information—but they usually retrieve different kinds of information.

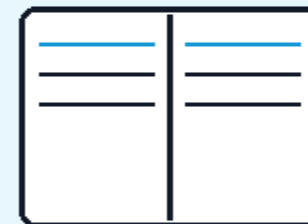
AGENT MEMORY



ABOUT THIS USER, TASK, OR EXPERIENCE

- ✓ Preferences and profile facts
- ✓ Past outcomes and progress
- ✓ Learned procedures or skills

RAG / KNOWLEDGE



ABOUT EXTERNAL DOCUMENTS OR KNOWLEDGE

- ✓ Policies and manuals
- ✓ Product documentation
- ✓ Research papers and knowledge bases

A strong agent may use both: memory for continuity, RAG for grounded knowledge.

WHERE MEMORY HELPS IN REAL LIFE

The same concept becomes useful in very different domains.



EDUCATION

Remembers mastered topics and recurring mistakes.



TRAVEL

Remembers budget, dietary needs, and past disruptions.



WORKPLACE

Remembers project decisions, owners, and unresolved tasks.



CUSTOMER SUPPORT

Remembers prior issues, attempted fixes, and preferences.



HEALTH ADMIN

Remembers appointment preferences—not medical truth without verification.



SOFTWARE

Remembers repository conventions and previous debugging outcomes.

The stored details should match the risk level of the domain.

MEMORY CAN HELP— OR QUIETLY HURT

A wrong memory can be worse than no memory because it feels familiar and trustworthy.



STALE

An old address or preference is treated as current.



FALSE

The agent stores an inference as if it were confirmed.



WRONG USER

Memory leaks across identities or accounts.



TOO MUCH

Noise crowds out the information needed now.



TOO SENSITIVE

Secrets or private details are stored unnecessarily.



UNDELETABLE

The user cannot inspect, correct, or remove it.

SAFE MEMORY: isolate users • verify facts • minimize storage • timestamp • allow correction and deletion

GOOD MEMORY IS SELECTIVE, RELEVANT & CORRECTABLE



WORKING

Keep the current task coherent.



SEMANTIC

Remember useful facts and preferences.



EPIODIC

Recall past sequences and outcomes.



PROCEDURAL

Apply approved rules and reusable skills.

NEXT

DAY 9: CONTEXT ENGINEERING

Giving the agent the right information at the right time

What is one thing you would want your personal agent to remember?