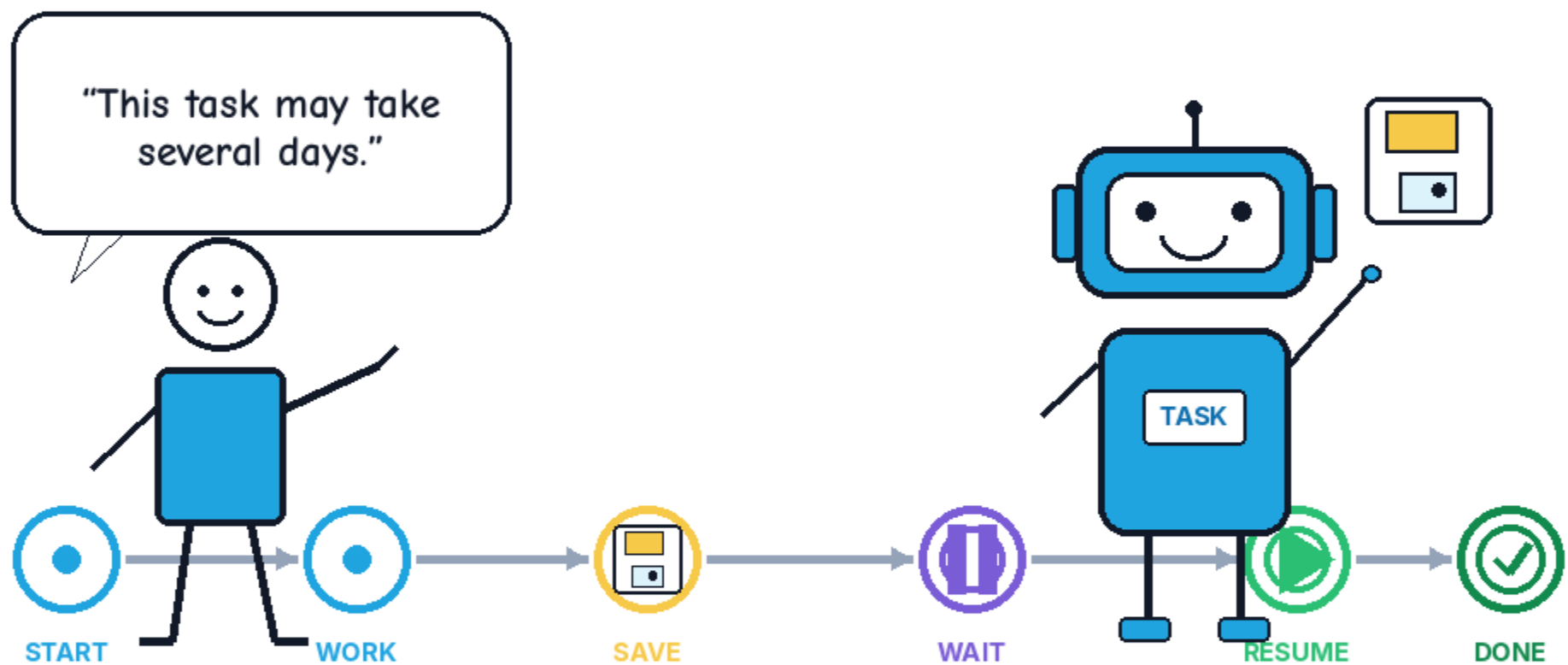


HOW DOES AN AI AGENT WORK FOR HOURS, DAYS... OR WEEKS?

Long-running agents, checkpoints and resumability — explained simply.



Long-running does not mean always running. It means continuing safely.

SOME TASKS DO NOT FINISH IN ONE CHAT

Real work often includes delays, approvals and changing information.



**EMPLOYEE
ONBOARDING**

Days or weeks



**INSURANCE
CLAIM**

Several days



**DEEP
RESEARCH**

Minutes or hours



**SOFTWARE
MIGRATION**

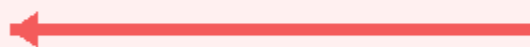
Hours or days

A long task often looks like: work → wait → continue later.

THINK OF A VIDEO GAME SAVE POINT

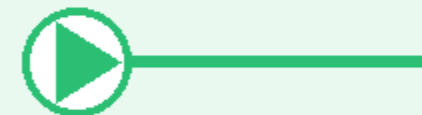
A checkpoint protects progress when something interrupts the task.

WITHOUT CHECKPOINTS



CRASH → BACK TO START

WITH CHECKPOINTS



RESUME FROM LAST SAVE

One interruption should not erase hours of completed work.

WHAT IS A LONG-RUNNING AGENT?

It keeps the task's identity and progress across time, failures, restarts and human pauses.

A system that can stop safely, remember where it was, and continue later without starting from zero.



20 MINUTES OR 2 WEEKS

Duration depends on the task — not the definition.



NOT ALWAYS "THINKING"

It can sleep while waiting for an event or a person.



SAME TASK, SAME IDENTITY

The system knows which job it is continuing.

Long-running agents alternate between active work and safe waiting.

MEMORY VS TASK STATE

They are related, but they answer different questions.

MEMORY

"What facts or preferences may help across tasks?"

Preference

Morning flights

Company rule

Refunds above ₹5,000 need approval

Past fact

The user already verified their address

TASK STATE

"Where is this specific task right now?"

Completed

Compared three flights

Waiting

Hotel approval pending

Remaining

₹35,000 budget and 2 retries

Memory supports future decisions. State lets this task continue.

THE THREE BUILDING BLOCKS

State changes. Checkpoints preserve it. Resume brings it back.



STATE

What the agent knows about the current task.



CHECKPOINT

A saved snapshot at a safe moment.



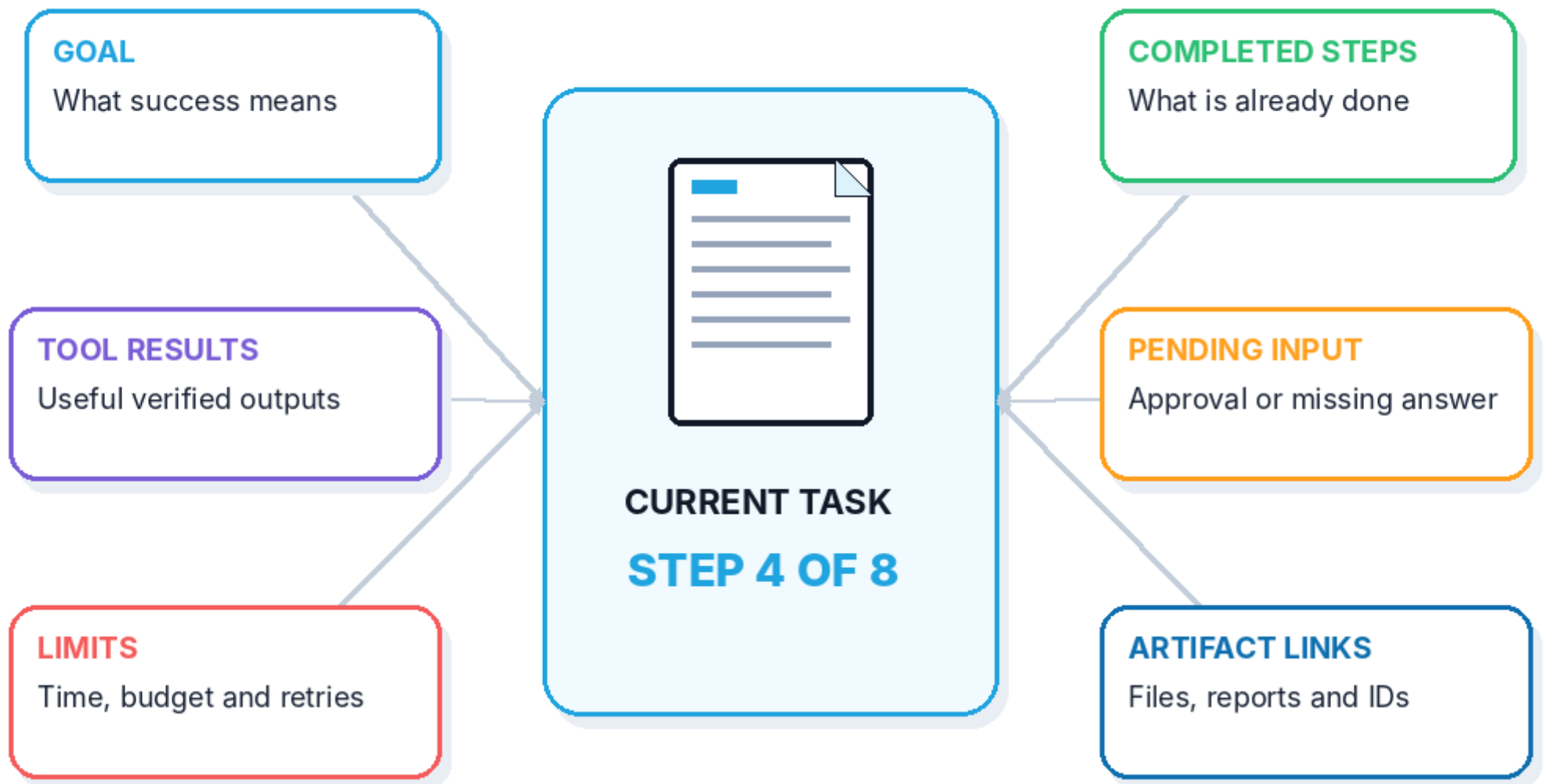
RESUME

Continue from the saved progress.

Together, these turn a fragile agent run into a durable task.

WHAT SHOULD THE AGENT SAVE?

Enough information to continue correctly — but not every raw log or secret.



Save references to sensitive data whenever possible — not the secrets themselves.

A CHECKPOINT IS A TASK SNAPSHOT

It records the state at a safe boundary so the run can be restored later.

CHECKPOINT · 10:32 AM

Task ID: ONB-204

Goal Prepare new employee for Day 1

Progress Step 4 of 8

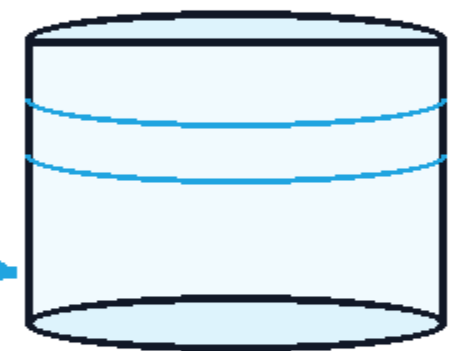
Done Forms sent; identity verified

Waiting Manager approval

Next Create accounts

Limits 1 of 3 retries used

Artifacts welcome_packet.pdf



DURABLE STORAGE

Database, object store or workflow runtime

SURVIVES RESTARTS

A checkpoint should live somewhere that survives app or server restarts.

THE DURABLE AGENT LIFECYCLE

A reliable task moves through work, save, wait, restore and continue.

The agent saves after meaningful progress.

A later worker can restore and continue.



Browser closes



Task continues



The interface may disappear. The task should not forget.

EXAMPLE: NEW EMPLOYEE ONBOARDING

One goal can involve several actions spread across many days.

GOAL Prepare a new employee for a smooth first day.



The agent pauses safely while waiting for the manager.

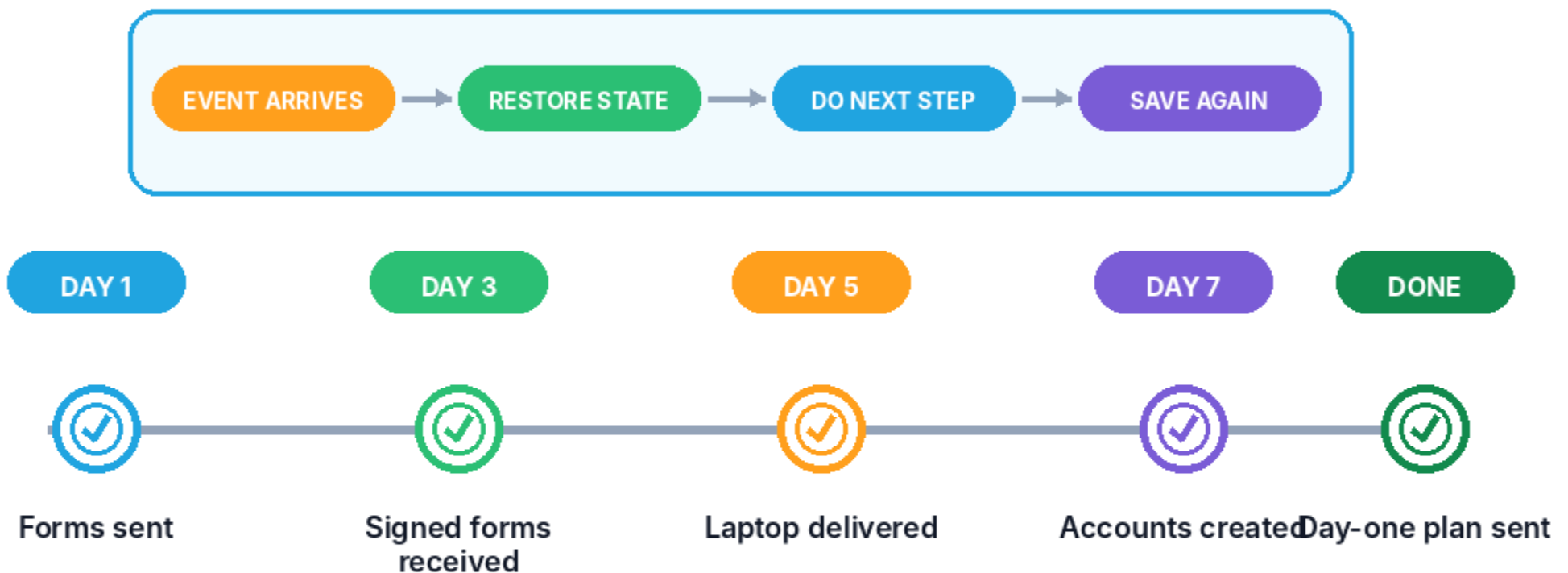


It does not keep a computer busy for days.

Checkpoint: forms sent, approval pending, next step known.

THE SAME AGENT, DAYS LATER

It wakes up only when an event arrives, restores the task and does the next useful step.

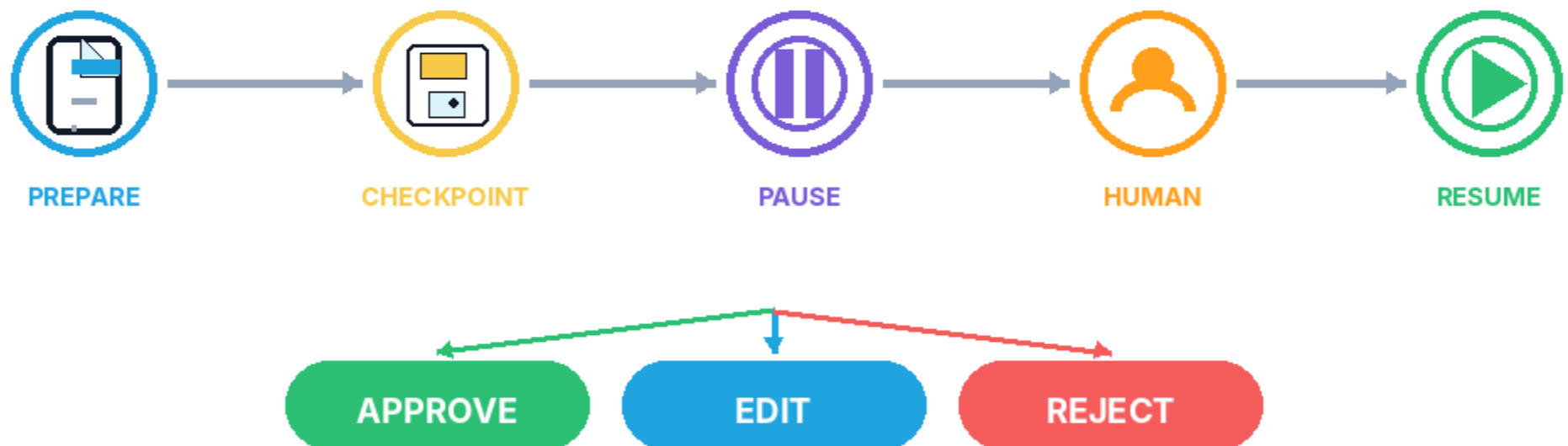


Long-running agents are often event-driven: work, sleep, wake and continue.

HUMAN APPROVAL WITHOUT LOSING PROGRESS

The agent can pause before a consequential action and resume after a clear decision.

"Create 25 new accounts and send welcome emails?"



Pause before sensitive actions. Continue only after an explicit response.

RETRY, RESUME OR RESTART?

These three recovery choices solve different problems.



RETRY

Repeat only the failed step.

Example: an API timed out.



RESUME

Restore a checkpoint and continue.

Example: the app or server restarted.



RESTART

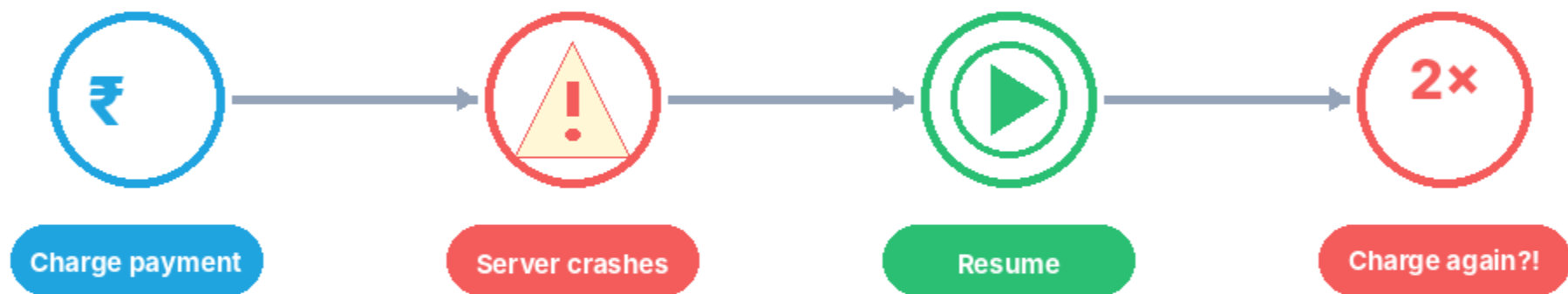
Begin the whole task again.

Use only when state is invalid or the goal changed.

Most failures should not erase completed work.

THE BIGGEST TRAP: DOING THE SAME ACTION TWICE

A crash can happen after the real-world action succeeds but before the agent records it.



MAKE THE ACTION SAFE TO REPEAT

- ✓ **Unique action key**: the same request cannot create a second payment.
- ✓ **Check before writing**: look for an existing result first.
- ✓ **Same request, same result**: return the earlier outcome instead of doing it again.

BUILDER WORD: IDEMPOTENT

Retries must never create duplicate payments, emails, bookings or records.

LONG TASKS CAN OVERFLOW THE CONTEXT

Keeping every message and tool result in the model's working context creates noise and cost.

BAD: DUMP IT ALL

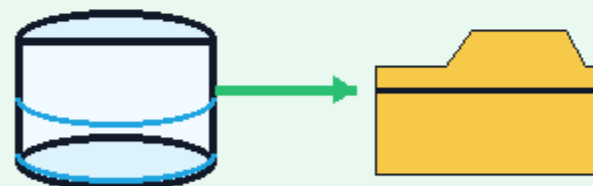


Huge context → repeated details, distraction and higher cost.

GOOD: COMPACT + RETRIEVE

TASK SUMMARY

Goal: onboarding
Done: 4 of 8 steps
Waiting: approval
Files: 3 linked artifacts



Retrieve old details only when needed.

Keep working context small. Keep durable progress outside it.

COMMON FAILURE MODES

Durability is more than saving state — the agent must resume correctly and safely.



NO CHECKPOINT

Hours of work disappear.



CHECKPOINT TOO LATE

Too much work repeats.



STALE STATE

The next action is wrong.



DUPLICATE SIDE EFFECT

Two charges, emails or bookings.



WAIT FOREVER

No timeout or escalation.



SAVE SECRETS CARELESSLY

Private data leaks.

A durable task must preserve progress, control retries and protect real-world actions.

PRODUCTION CHECKLIST

Build for interruptions on purpose — not as an afterthought.

- ✓ Store task state in durable storage
- ✓ Checkpoint after meaningful steps
- ✓ Use safe retries and unique action keys
- ✓ Set timeouts, retry limits and escalation rules
- ✓ Add human approval for consequential actions
- ✓ Show status updates and keep an audit trail
- ✓ Define completion, cancellation and failure states
- ✓ Test crash-and-resume scenarios deliberately



TEST: STOP THE AGENT MID-TASK



A reliable system should recover from the interruption you intentionally create.

DAY 15 TAKEAWAY

Long-running does not mean never stopping.

A durable agent can stop safely, remember exactly where it was, and continue without repeating harmful actions.

STATE

+

CHECKPOINTS

+

SAFE RESUME

+

HUMAN APPROVAL



NEXT UP

DAY 16 — MCP: ONE STANDARD WAY TO CONNECT AGENTS TO TOOLS AND DATA

What task in your life or work takes days and could benefit from safe resume?