

The green check can hide a lost state.

The same correct file can come from a true resume, a clean restart or a quiet human rescue.

GREEN CHECK

Output accepted

Required file exists

Acceptance test passes

Destination bytes match

RECOVERY PROOF

Trajectory joined

Fault detected

Checkpoint identified

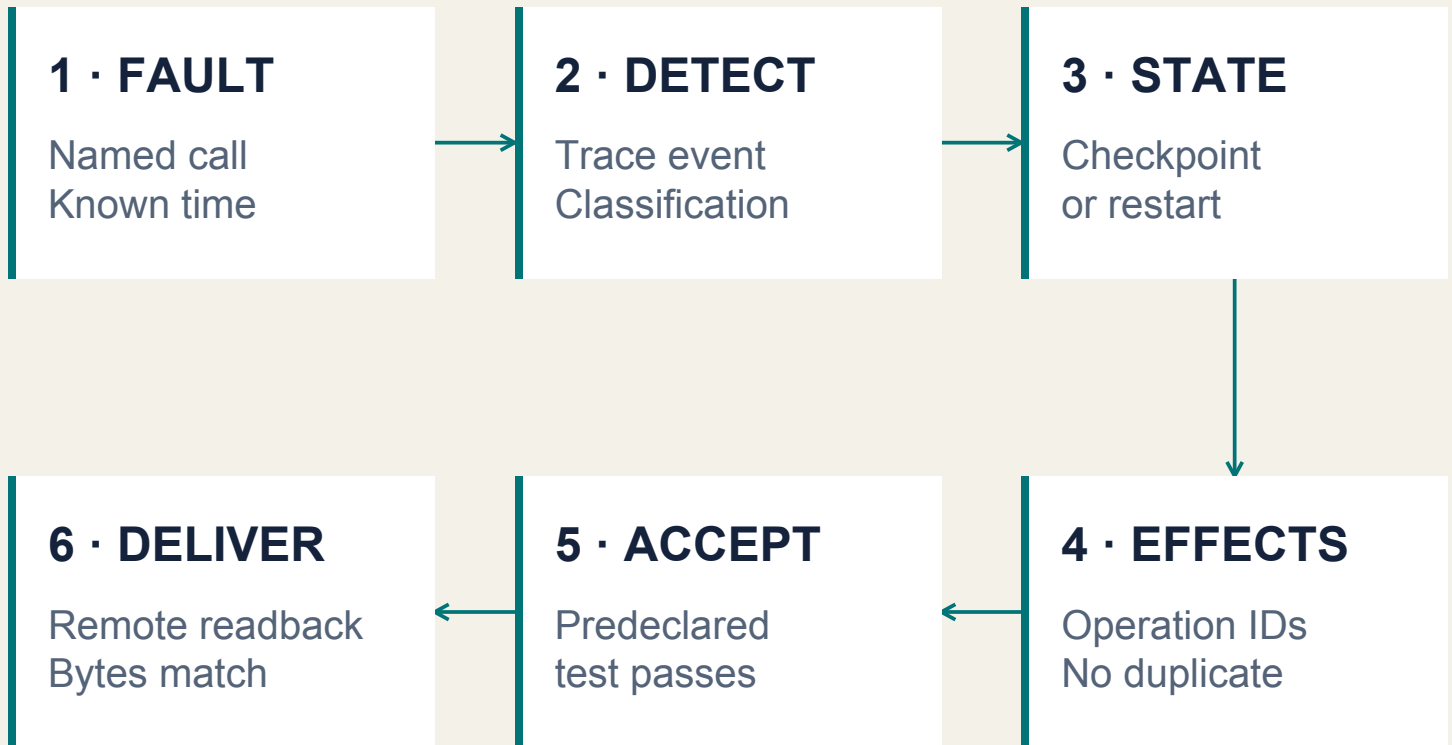
No duplicate effects

The source Special proposes the test. No recovery experiment was run.

Acceptance verifies the destination. Recovery evidence verifies the path.

Six receipts make one evidence spine.

A missing join leaves resume, restart and rescue indistinguishable.



The snake path preserves event order: fault → detect → state → effects → acceptance → delivery.

The final file is the last node—not the whole proof.

Restartability is not stateful recovery.

Both can generate identical final bytes. Only one preserves previously established work.

CLEAN RESTART

Restartable

Discard earlier state

Repeat completed work

May still pass acceptance

CHECKPOINT RESUME

State preserved

Name trusted checkpoint

Continue from boundary

Bind path to artifact

If a person reconstructs the state, label the result accepted with human rescue—not autonomous recovery.

Name the mechanism that actually happened.

Give recovery its own clock.

Run one fixed read-only task as twins: control intact, treatment with one predeclared timeout.

DIMENSION	CONTROL	TREATMENT – CONTROL
Wall time	Start → delivery	Observed delay
Calls + tokens	Same task budget	Repeated work
Direct cost	Same price basis	Observed increment
Human minutes	Routine review	Rescue kept separate

Recovery tax is a vector, not one blended score. No numerical result is claimed.

Subtract like from like. Keep the dimensions separate.

Freeze the twin before the fault.

A kinder prompt, extra retries or a whispered hint changes the system under test.

- 01 Task**
Same fixtures + acceptance rule
- 02 System**
Same model, prompt, harness + tools
- 03 Budget**
Same calls, retries, time + cost ceiling
- 04 Fault**
One named timeout at one fixed call
- 05 Safety**
Read-only first; stop + rollback ready

Repeat across fault positions and seeds before making a statistical recovery-rate claim.

One timeout. One causal change. One bounded claim.

The harness can own the bill.

In one reported case, agents repeatedly polled a long backtest and filled context with status checks.

BEFORE

~3 hours

Repeated polling

Token waste

Context pollution

AFTER

~3–5 minutes

Caching change

Memory packing

Same reported workload

Company-reported example. It is not an independent benchmark or a universal agent result.

Memory and waiting policy can dominate the cost of persistence.

Verify the path— not only the finish.

The destination matters. The evidence spine tells you whether the system knew how it got there.

FINAL SIGNAL

Output accepted

Checkpoint + operation IDs + artifact hash + readback

Call a restart restartability, a rescue human help, and a verified checkpoint path one accepted autonomous recovery.

Protocol published · Experiment not run · No production fault injected

A green check is the last receipt, not all six.