

066: Replay & Restart Modes (Forced Tool Replay vs Harness)

Date: January 15, 2026
Status: Draft (Discovery Notes)

1. Overview

The Universal Agent supports two distinct restart mechanisms:

- 1. **Forced Tool Replay (Crash Recovery)**
- 2. **Harness Restart (Ralph loop for long-running tasks)**

These modes serve different purposes and are triggered by different state transitions.

2. Forced Tool Replay (Crash Recovery)

This mode is triggered when a run resumes with in-flight tool calls (*prepared or running*).

Process summary:

- Load in-flight tool calls from the ledger.
- Promote pending receipts if possible (skip replay).
- If replay is required, enqueue tool calls into *forced_tool_queue*.
- Enter *forced_tool_mode_active* and run a strict replay prompt.
- If replay cannot complete, mark run *waiting_for_human*.

This flow is handled by *reconcile_inflight_tools()* and *_build_forced_tool_prompt()*.

3. Harness Restart (Long-Running)

Harness restarts occur when:

- The agent outputs the completion promise **incorrectly or not at all**, or
- The harness verification fails (missing/invalid artifacts), or
- Context exhaustion triggers a controlled reset.

On restart, the harness injects *mission.json* + *mission_progress.txt* into the next prompt and increments *iteration_count*.

4. Key Distinctions

Feature	Forced Tool Replay	Harness Restart
-----	-----	-----
Purpose	Crash recovery	Long-running tasks
Trigger	In-flight tools	Promise/verification/context

Context reset	Minimal	Explicit (client history cleared)
Prompt style	Strict replay prompt	Mission resume prompt
Status update	waiting_for_human if stuck	restart until promise satisfied

5. Related Files

- *main.py* (reconcile_inflight_tools, on_agent_stop, forced_tool_queue)
- *durable/ledger.py* (tool call persistence)
- *durable/state.py* (run status + iteration counts)