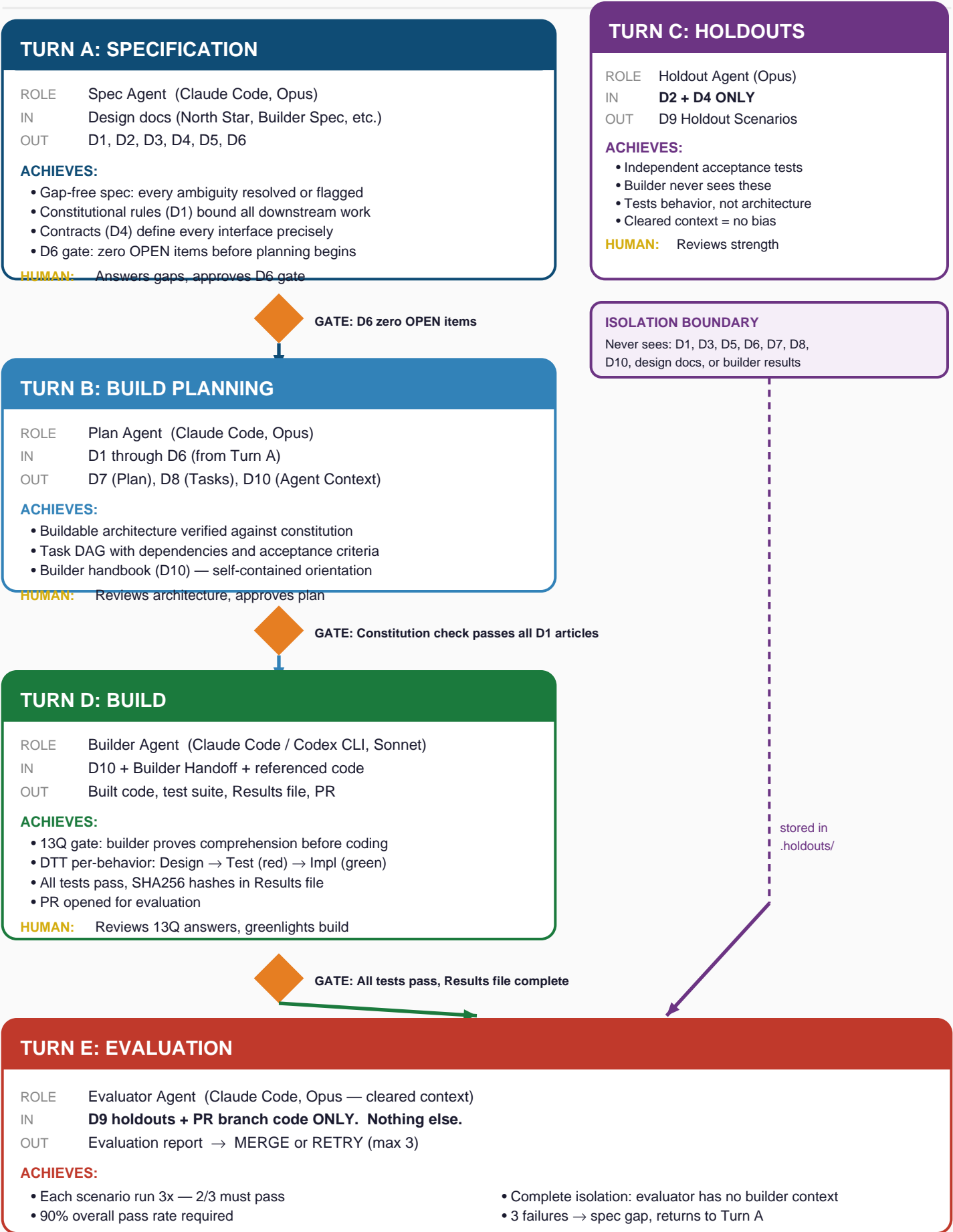


The Sawmill Pipeline

A Product-Agnostic Dark Factory — L3 Maturity — 5 Turns, Spec to Verification



What Each Turn Achieves

Summary of outcomes, gates, and the verification chain

Turn A SPECIFICATION

Eliminates ambiguity before it reaches a builder

- Every boundary walked. Every gap flagged or resolved.
- Constitutional rules established — immutable constraints.
- Contracts precisely define every input, output, side-effect, error.
- Controlled redundancy seeded across D1-D6 for drift detection.
- D6 gate: zero OPEN items. Nothing unresolved enters planning.

Turn B BUILD PLANNING

Translates 'what to build' into 'how to build it' — verified against the spec

- Architecture verified against every D1 constitutional article.
- Tasks decomposed with dependency DAG and acceptance criteria.
- Every D2 scenario and D4 contract has at least one task.
- Builder handbook (D10) enables cold-start orientation.
- Cross-document checks catch drift between spec and plan.

Turn C HOLDOUT AUTHORIZING

Creates the independent quality gate the builder cannot game

- Behavioral tests written from D2+D4 only — no implementation knowledge.
- Cleared agent context prevents information leakage.
- Executable scenarios: setup/execute/verify/cleanup in bash.
- Coverage matrix ensures all P0/P1 scenarios are tested.

Turn D BUILD

Produces verified code through disciplined, per-behavior TDD

- 13Q gate proves comprehension before a single line is written.
- DTT loop: Design → Test (must fail) → Implement (minimum) → Refactor.
- Builder works from D10 + handoff only — constrained scope.
- Results file with SHA256 hashes provides audit trail.

Turn E EVALUATION

Validates code against tests the builder never saw

- Evaluator has cleared context — no builder reasoning, no spec.
- Each scenario runs 3x for statistical confidence.
- 90% pass rate gate. Failure traces to spec gaps, not code bugs.
- 3 retries then escalate — systematic, not heroic.

THE CHAIN OF VERIFICATION

Each turn's output is verified against a DIFFERENT document than the one it was written from.

D7 checks against D1 (not D2). D8 traces to D2+D4. Holdouts test D2+D4 without seeing D7-D10. Evaluator sees only D9+code.