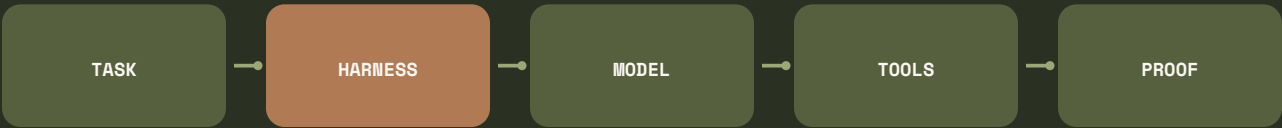


MODEL-AGNOSTIC / MECHANISM-FIRST

Build Your Own Agentic Coding Harness

A practical control plane for state, tools, permissions, context, execution, verification, and proof - with the model kept replaceable.



- STATE OWNED
- TOOLS TYPED
- FAILURE VISIBLE
- PROOF ATTACHED

01 / THE BOUNDARY

The model is replaceable. The harness is the product.

Put orchestration, permissions, execution, and verification outside the model so a provider change does not rewrite the system.

COMPATIBILITY TRUTH

Support any model that reliably returns text. Prefer native tools; use a constrained action envelope only as a degraded fallback. Model independence is not capability parity.

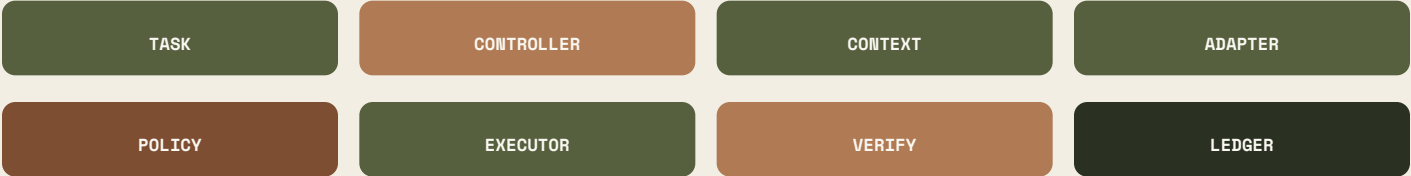
MODEL OWNS

- Generate candidate text
- Suggest tool calls
- Interpret observations
- Propose code changes
- Choose a likely next move

HARNESS OWNS

- Run state and budgets
- Context and provenance
- Tool schemas and policies
- Workspace and sandbox
- Verification, trace, and replay

REFERENCE FLOW



PRIMARY SOURCES

[OPENAI TOOLS](#) [ANTHROPIC TOOLS](#) [GEMINI TOOLS](#)

Normalize the edge, expose the differences.

The common contract carries messages, tools, streaming events, usage, and errors. Capability flags prevent lowest-common-denominator design.

CANONICAL INTERFACE

```
interface ModelAdapter {
  id: string;
  capabilities(): Promise<Capabilities>;
  generate(
    request: ModelRequest,
    signal: AbortSignal
  ): AsyncIterable<ModelEvent>;
}

ModelEvent =
  text_delta | tool_call | usage |
  done | error
```

CAPABILITY NEGOTIATION

NATIVE TOOLS	YES / NO
STRUCTURED	STRICT / BEST / NONE
PARALLEL	YES / NO
STREAMING	YES / NO
CONTEXT	TOKEN LIMIT

ADAPTER RULES

- Preserve tool-call IDs
- Retry transport, not side effects
- Record resolved model version
- Keep provider options at the edge

NATIVE TOOL CYCLE



TEXT-ONLY FALLBACK

Require one JSON action envelope. Parse, schema-check, and reject anything outside it. Never execute prose that only resembles a command.

PRIMARY SOURCES

- [OPENAI FLOW](#) [ANTHROPIC TOOL USE](#) [GEMINI FUNCTION CALLING](#)

03 / RUN CONTROLLER

The model proposes. The state machine decides.

Persist every transition before the next side effect. A crash should resume from a known state, not infer what may have happened.

STATE MACHINE



PERSIST THIS RUN RECORD

- Task + acceptance criteria
- Workspace root + revision
- Adapter + requested model
- Step, time, token, cost budgets
- Approval decisions + scope
- Current status + timestamps

STOP CONDITIONS IN CODE

- Verifier accepts completion
- Hard budget is exhausted
- User cancels or denies approval
- Failure repeats past retry policy
- No-progress detector fires
- Policy marks the run terminal

EXACTLY-ONCE EFFECT RULE

Assign every proposed tool call an idempotency key. Retry model transport failures freely. Retry a side effect only when the executor can prove the earlier attempt did not commit.

Terminal states: COMPLETED / FAILED / CANCELLED / BUDGET_EXHAUSTED

Typed capabilities first. Execution second.

Every action crosses schema validation and policy review before it reaches a workspace, process, network, or external system.

START WITH SEVEN TOOLS

workspace.list	READ
workspace.read_file	READ
workspace.search	READ
workspace.apply_patch	WRITE
workspace.diff	READ
shell.run	EXECUTE
verify.run	EXECUTE

POLICY OUTCOMES

ALLOW

Safe inside declared scope

APPROVE

Consequential or external

DENY

Forbidden by boundary

APPROVAL MUST BIND

tool + normalized args + scope + expiry + user

DEFAULT BOUNDARY

- Workspace reads: allow
- In-root patch: allow, then show diff
- Named tests: allow with timeout
- Network or dependency install: require approval
- Secrets, out-of-root writes, destructive effects: deny

PRIMARY SOURCES

[JSON SCHEMA](#) [MCP TOOLS](#) [OWASP THREATS](#)

Constrain the environment. Budget the prompt.

Generated commands and retrieved files are untrusted data. Isolation limits impact; provenance keeps instructions separate from evidence.

EXECUTION BOUNDARY

- One explicit workspace root
- Clean snapshot before writes
- Read-only except declared roots
- No secrets or network by default
- Non-root user + resource limits
- Timeout + process-tree cleanup
- Bounded stdout and stderr
- Diff after every write phase

CONTEXT LAYERS

POLICY	NEVER SUMMARIZE
TASK	NEVER SUMMARIZE
REPOSITORY	REFRESH ON CHANGE
WORKING STATE	KEEP EXPLICIT
RECENT EVIDENCE	TRIM BY RELEVANCE
HISTORY	SUMMARIZE WITH RECEIPTS

PATCH FLOW

READ + HASH

MODEL PATCH

HASH CHECK

APPLY IN ROOT

RETURN DIFF

VERIFY

A stale file fails closed. The harness never rewrites a newer version blindly.

PRIMARY SOURCES

[OWASP SECURE AGENTS](#) [NIST GENAI PROFILE](#)

A completion claim is not proof.

Record every control decision, then verify with deterministic checks before using a model grader for anything subjective.

APPEND-ONLY RUN LEDGER



VERIFICATION PIPELINE

- Diff stays inside requested scope
- Required files and artifacts exist
- Format, lint, types, tests, build pass
- Secret and policy scans do not regress
- Acceptance criteria map to evidence
- Workspace remains reviewable

REPLAY MODES

DETERMINISTIC

Reuse stored model and tool outputs to reproduce controller decisions.

LIVE

Rerun the same fixture with another adapter and compare outcomes.

MINIMUM METRICS



PRIMARY SOURCES

[OPENTELEMETRY GENAI](#) [NIST GENAI PROFILE](#)

Ship the control plane in five increments.

Start read-only. Add writes, execution, recovery, and provider comparisons only after each previous boundary has proof.

01

READ-ONLY

Adapter +
list/read/search + ledger

Every answer cites
files

02

PATCH

Structured patch + roots
+ hashes

Stale patch is rejected

03

EXECUTE

Named checks +
sandbox + approvals

Network cannot run
silently

04

RESUME

Persisted state + pause
+ replay

Crash resumes exactly
once

05

MATRIX

Three adapters + shared
fixture suite

One config swaps the
model

SYSTEM PROOF / DEFINITION OF DONE

☐ Swap the model through configuration

☐ Expose capability differences

☐ Validate every tool input

☐ Keep every write inside roots

☐ Bound every command and process

☐ Gate network and destructive effects

☐ Pause and resume approvals once

☐ Reject false completion with evidence

TASK DEFINED

MECHANISM VISIBLE

FAILURE HANDLED

REUSABLE TOMORROW

PRIMARY SOURCES

[OPENAI](#) [ANTHROPIC](#) [GEMINI](#) [MCP](#) [JSON SCHEMA](#) [OPENTELEMETRY](#) [OWASP](#) [NIST](#)