

John Fail-Closed Specs

Current reconstructed John replacement/drift-prevention specs and review packet

- [John Agent Replacement Plan — Current Reconstructed](#)
- [John Drift Prevention Spec — Current Reconstructed](#)
- [John Fail-Closed CEO Review Packet — 2026-05-23](#)
- [John Blueprint Cost-Waiver Preflight — 2026-05-23](#)

John Agent Replacement Plan — Current Reconstructed

John Replacement Plan — Reconstructed Current Plan

Status: **CURRENT_RECONSTRUCTED_PLAN_PENDING_CEO_SIGNOFF**

Date: 2026-05-23

MC: #10599

Replaces: tombstone/stale marker created 2026-05-23 after the original file was absent.

Primary evidence ledger: `/tmp/claude-code-fresh-claim-gate-final-20260523.md`

0. Executive decision

Do **not** replace John by adding more advisory agents. Replace the unsafe behavior with deterministic, fail-closed enforcement at every output/delegation boundary:

1. Claude Code Stop hooks.
2. Claude Code PreToolUse delegation hooks.
3. Shared claim classifier.
4. Pi extension output boundary.
5. Virtual-company `agent-runner.js` response boundary.
6. Company Mesh response boundary.
7. Evidence-path and cost gates before large workflows.

Memory, HiveMind, RAG snippets, old state, and peer recollection are advisory only and must never be treated as evidence for ALAI/MC/system-state claims.

1. Current implemented foundation

Claude Code boundary

Current `/Users/makinja/.claude/settings.json` includes these enforcement hooks:

- `PreToolUse Task|WebSearch|WebFetch`: `bash ~/.claude/hooks/pre-action-da-gate.sh`
- `Stop`: `bash ~/.claude/hooks/alai-claim-gate.sh`
- `Stop`: `python3 ~/.claude/hooks/john-determinism-gate.py`
- `Stop`: `python3 ~/.claude/hooks/claim-auto-probe-gate.py`
- `UserPromptSubmit`: `bash ~/.claude/hooks/boot-enforcer.sh`

Current wrapper behavior:

- `/Users/makinja/.claude/hooks/alai-claim-gate.sh` runs `/Users/makinja/system/tools/alai-claim-gate.js` on Claude Code transcripts.
- It now fails closed with `CLAUDE_STOP_HOOK_MISSING_TRANSCRIPT` if Stop hook payload has no readable transcript.

Shared claim gate

`/Users/makinja/system/tools/alai-claim-gate.js` blocks factual/system-state claims without evidence. Current violation anchors include:

- `STATE_CLAIM_WITHOUT_EXISTING_EVIDENCE_PATH`
- `ALAI_FACTUAL_CLAIM_WITH_ZERO_TOOL_CALLS`

Pi boundary

- `/Users/makinja/.pi/agent/extensions/alai-claim-gate.ts` defaults `ALAI_CLAIM_GATE_MODE` to `hard`.
- `/Users/makinja/.pi/agent/extensions/company-mesh-tools.ts` explicitly states advisory sources are `ADVISORY_NOT_EVIDENCE`.

Virtual-company boundary

- `/Users/makinja/system/tools/agent-runner.js` runs shared claim gate before printing/saving agent output.
- `/Users/makinja/system/tools/company-mesh.js` runs shared claim gate before DB insertion for mesh responses.

2. Evidence already obtained

Evidence artifacts:

- `/tmp/alai-hardening-evidence-20260523.md`
- `/tmp/alai-claim-gate-deadlock-fix-20260523.md`
- `/tmp/alai-fail-closed-retest-20260523.md`

- /tmp/pi-virtual-company-claim-gate-20260523.md
- /tmp/pi-claim-gate-extension-harness-20260523.md
- /tmp/pi-fresh-session-claim-gate-20260523.md
- /tmp/agent-runner-claim-gate-smoke-20260523.md
- /tmp/pi-virtual-company-advisory-contract-20260523.md
- /tmp/smoke-test-agent-and-dev-state-cleanup-20260523.md
- /tmp/john-specs-stale-evidence-20260523.json
- /tmp/john-missing-specs-stale-markers-20260523.md
- /tmp/claude-code-fresh-claim-gate-final-20260523.md

Key fresh Claude Code evidence:

- Fresh normal-session hallucination smoke produced the unsupported sentence `The MC task is completed and blueprint MUST can start.`
- Claude Code Stop hook blocked it with exit code `2`.
- Shared claim gate violations were `STATE_CLAIM_WITHOUT_EXISTING_EVIDENCE_PATH` and `ALAI_FACTUAL_CLAIM_WITH_ZERO_TOOL_CALLS`.
- `--no-session-persistence` no longer bypasses the claim gate; missing transcript fails closed.
- Synthetic readable-transcript regression: no-evidence blocks with `rc=2`, evidence-path retry allows with `rc=0`.

3. Replacement architecture

3.1 John core behavior

John may answer factual ALAI/MC/system-state questions only after tool verification. If current evidence is absent, John must answer one of:

- `I have not verified that yet.`
- `BLOCKED: needs current tool evidence.`
- `I can verify with <specific tool/path> if you approve.`

John must not claim:

- task completion,
- MC completion,
- blueprint readiness,
- hook activation,
- deployment/live status,
- agent execution,
- evidence existence,

unless a same-turn tool or cited existing evidence path supports it.

3.2 Enforcement-first design

The replacement is not a persona rewrite. It is a boundary system:

1. **Prompt intake:** boot/checklist freshness gate.
2. **Tool dispatch:** delegation cannot proceed without MC reference.
3. **Assistant final output:** Claude Stop hooks block unsupported claims.
4. **Pi final output:** Pi extension blocks unsupported claims hard by default.
5. **Agent output:** `agent-runner.js` blocks before response is saved/printed.
6. **Mesh output:** `company-mesh.js` blocks before DB write.
7. **Evidence retry:** existing evidence path can allow claims when the path exists.

3.3 Advisory-source quarantine

Every prompt or worker context must include this contract:

“Memory, HiveMind, RAG snippets, old state, and peer recollection are ADVISORY_NOT_EVIDENCE for ALAI, MC, deployment, hook, workflow, agent, production, or task-status claims.

4. Blueprint MUST gate

Blueprint MUST workflows may start only if all conditions are true:

1. Fresh Claude Code claim-gate smoke has passed.
2. Missing-transcript/no-session bypass is fail-closed.
3. Pi and virtual-company output gates are hard or explicitly waived.
4. Cost review has been done for the current day/session.
5. User explicitly approves the run or provides a written waiver.
6. The workflow is run through a wrapper/checklist, not free chat.

Current state as of this reconstruction:

- Conditions 1 and 2 have evidence in `/tmp/claude-code-fresh-claim-gate-final-20260523.md`.
- Pi/virtual-company evidence exists in the listed `/tmp` artifacts.
- Cost is high today: latest observed cost probe returned `$45.4829` total for Claude CLI usage.
- Therefore, large paid blueprint MUST execution still requires explicit approval/waiver.

5. Implementation phases

Phase A — Completed hardening baseline

- Fail-closed Claude hooks.
- Shared claim gate deadlock fix.
- Claude fresh-session smoke.
- Pi hard default.
- Agent-runner shared output gate.
- Company Mesh shared output gate.
- Dedicated smoke-test identity.
- Operational `dev` state cleanup.
- Stale missing John specs marked and then reconstructed.

Phase B — Immediate next local work

1. Validate syntax for modified code.
2. Validate these reconstructed specs exist and are not tombstones.
3. Create an evidence artifact for the reconstruction.
4. Do **not** mark MC #10599 or #10570 complete without CEO sign-off and any required commit/indexing evidence.

Phase C — Optional commit/index/sign-off work

Only after approval:

1. Commit or otherwise persist changed source files.
2. Index summary into approved memory mechanism if required.
3. Update MC #10599/#10570 status with evidence paths.
4. Run blueprint MUST wrapper/checklist if cost approval exists.

6. Risk controls

- Break-glass for missing transcript exists only via `ALAI_CLAIM_GATE_ALLOW_MISSING_TRANSCRIPT=1` and must be treated as explicit maintenance waiver.
- Claude hook safe mode must not disable claim gates silently.
- Smoke tests must use dedicated smoke identity, not operational `dev` state.
- Any future stale/missing path must be tombstoned before it is reconstructed.

7. Open acceptance items

- CEO sign-off is pending.
- Commit/indexing evidence is pending.
- MC #10599 should remain open until sign-off and persistence requirements are satisfied.
- Blueprint MUST execution is still blocked on cost/approval despite gate readiness evidence.

John Drift Prevention Spec — Current Reconstructed

John Drift-Prevention Spec — Reconstructed Current Spec

Status: **CURRENT_RECONSTRUCTED_SPEC_PENDING_CEO_SIGNOFF**

Date: 2026-05-23

MC: #10570

Replaces: tombstone/stale marker created 2026-05-23 after the original file was absent.

Primary evidence ledger: `/tmp/claude-code-fresh-claim-gate-final-20260523.md`

0. Scope and authority

This is a reconstructed current spec from deterministic files and smoke tests. It is **not** the missing original 2026-05-02 spec.

Authority hierarchy for John/ALAI operational claims:

1. Current tool output and existing evidence artifacts.
2. Current source files in `/Users/makinja/.claude`, `/Users/makinja/.pi/agent`, and `/Users/makinja/system`.
3. MC task state from `node /Users/makinja/system/tools/mc.js show <id>`.
4. Memory, HiveMind, RAG snippets, context bundles, and old agent state:

ADVISORY_NOT_EVIDENCE.

No response may treat advisory sources as proof of implementation, deployment, MC completion, blueprint readiness, or hook activation.

1. Memory-mistrust protocol

Rule

John must not convert memory feedback or historical context into current-state claims without a same-turn deterministic check.

Required verification by claim type

Claim type	Required evidence
MC status / owner / priority	<code>node /Users/makinja/system/tools/mc.js show <id></code> or <code>list</code> output
Hook active/wired	<code>/Users/makinja/.claude/settings.json</code> and executable hook file exists
Hook behavior	synthetic or fresh-session smoke result with <code>rc</code> / hook event evidence
Pi extension active	Pi settings/extension source plus fresh Pi or extension harness smoke
Virtual-company / mesh response safety	<code>agent-runner.js</code> / <code>company-mesh.js</code> shared gate source plus smoke evidence
Blueprint/MUST readiness	current blueprint path plus gate/test evidence; not memory text
Production/deploy state	live health/log/browser evidence

Deterministic implementation anchors

Current evidence shows these active Claude hooks in `/Users/makinja/.claude/settings.json`:

- `PreToolUse Task|WebSearch|WebFetch`: `bash ~/.claude/hooks/pre-action-da-gate.sh`
- `Stop`: `bash ~/.claude/hooks/alai-claim-gate.sh`
- `Stop`: `python3 ~/.claude/hooks/john-determinism-gate.py`
- `Stop`: `python3 ~/.claude/hooks/claim-auto-probe-gate.py`
- `UserPromptSubmit`: `bash ~/.claude/hooks/boot-enforcer.sh`

Evidence command output for this wiring was captured during 2026-05-23 reconstruction.

2. One CEO sentence = one bounded action

A single CEO instruction must not be expanded into an unbounded multi-agent tree.

Allowed immediately

- Read current files.
- Run narrow probes.
- Patch small deterministic gates.
- Write an evidence artifact.
- Ask for explicit approval when cost/risk exceeds threshold.

Requires explicit escalation before dispatch

- Creating MC EPICs.
- Dispatching multiple teams.
- Running paid/fresh-model smoke tests when daily spend is high.
- Starting blueprint MUST or large validation workflows.
- Production deploys or destructive cleanup.

Escalation contract

Before escalation John must state:

1. The exact requested action.
2. The deterministic premise already verified.
3. Estimated cost/risk.
4. Required approval or waiver.
5. Evidence path where results will be written.

3. MC EPIC creation preconditions

Before creating or modifying MC EPICs, John must verify:

1. The CEO request is current and not a stale memory replay.
2. The referenced path/task exists now.
3. The requested work cannot be completed as a direct local patch.
4. Cost is acceptable or approval exists.
5. The target owner/company route is source-of-truth verified.
6. Acceptance criteria are measurable with evidence artifacts.

If any precondition fails, John must stop and report `BLOCKED` or `NEEDS_INPUT`, not create recursive work.

4. Bash/tool enforcement points

Prompt discipline is insufficient. Drift prevention must be enforced at these boundaries:

4.1 Claude Code prompt/session boundary

- `/Users/makinja/.claude/hooks/boot-enforcer.sh` blocks stale boot/checklist state via exit 2.
- `/Users/makinja/.claude/hooks/alai-claim-gate.sh` invokes shared claim gate.
- `/Users/makinja/.claude/hooks/alai-claim-gate.sh` now fails closed if `transcript_path` is missing/unreadable, emitting `CLAUDE_STOP_HOOK_MISSING_TRANSCRIPT`.
- `/Users/makinja/.claude/hooks/john-determinism-gate.py` blocks AI OS / John / blueprint / MC execution claims without same-turn tool evidence.
- `/Users/makinja/.claude/hooks/claim-auto-probe-gate.py` is hard by default and transcript-aware.

4.2 Delegation boundary

- `/Users/makinja/.claude/hooks/pre-action-da-gate.sh` blocks Task delegation without MC reference.

4.3 Shared claim classifier

- `/Users/makinja/system/tools/alai-claim-gate.js` emits violations including:
 - `STATE_CLAIM_WITHOUT_EXISTING_EVIDENCE_PATH`
 - `ALAI_FACTUAL_CLAIM_WITH_ZERO_TOOL_CALLS`

4.4 Pi boundary

- `/Users/makinja/.pi/agent/extensions/alai-claim-gate.ts` defaults `ALAI_CLAIM_GATE_MODE` to `hard`.
- `/Users/makinja/.pi/agent/extensions/company-mesh-tools.ts` states Memory/HiveMind/RAG/old state/peer recollection are `ADVISORY_NOT_EVIDENCE`.

4.5 Virtual-company / mesh boundary

- `/Users/makinja/system/tools/agent-runner.js` runs the shared claim gate before printing/saving agent responses.
- `/Users/makinja/system/tools/company-mesh.js` runs the shared claim gate before writing mesh responses to DB.

5. Anti-pattern catalog

5.1 Petter T6 mis-diagnosis pattern

Bad pattern: infer a technical diagnosis from prior context and dispatch remediation without reading current files/logs.

Required behavior: verify source files/logs first, produce one bounded finding, then ask before broader dispatch.

5.2 AWS phantom drift pattern

Bad pattern: treat memory or old infra assumptions as proof that AWS resources/configuration exist.

Required behavior: use current infra source-of-truth tools or cloud CLI evidence before saying any resource exists, is broken, or was fixed.

5.3 Drift-after-step1-completion pattern

Bad pattern: after completing one narrow fix, invent a larger workflow and continue without approval.

Required behavior: stop after the bounded action, write evidence, and ask for approval before the next phase.

6. Validation evidence from 2026-05-23

Evidence artifacts:

- /tmp/alai-hardening-evidence-20260523.md
- /tmp/alai-claim-gate-deadlock-fix-20260523.md
- /tmp/alai-fail-closed-retest-20260523.md
- /tmp/pi-virtual-company-claim-gate-20260523.md
- /tmp/pi-claim-gate-extension-harness-20260523.md
- /tmp/pi-fresh-session-claim-gate-20260523.md
- /tmp/agent-runner-claim-gate-smoke-20260523.md
- /tmp/smoke-test-agent-and-dev-state-cleanup-20260523.md
- /tmp/john-specs-stale-evidence-20260523.json
- /tmp/john-missing-specs-stale-markers-20260523.md
- /tmp/claude-code-fresh-claim-gate-final-20260523.md

Fresh Claude Code evidence from /tmp/claude-code-fresh-claim-gate-final-20260523.md :

- Normal-session hallucination smoke blocked the claim The MC task is completed and blueprint MUST can start.
- Stop hook exit code was 2.

- Violations were `STATE_CLAIM_WITHOUT_EXISTING_EVIDENCE_PATH` and `ALAI_FACTUAL_CLAIM_WITH_ZERO_TOOL_CALLS`.
- `--no-session-persistence` missing-transcript bypass is patched to fail closed.
- Readable-transcript wrapper regression: no-evidence `rc=2`, existing evidence path `rc=0`.

7. Open acceptance items

- CEO sign-off is pending.
- This spec has not been committed by this document alone.
- MC #10570 should not be marked complete until sign-off/commit/indexing requirements are explicitly satisfied with evidence.

John Fail-Closed CEO Review Packet — 2026-05-23

CEO review packet — John fail-closed reconstruction — 2026-05-23

Decision requested

Please review and either sign off or request changes for these reconstructed current specs:

1. `/Users/makinja/system/specs/john-replacement-plan.md`
 - Status inside document: `CURRENT_RECONSTRUCTED_PLAN_PENDING_CEO_SIGNOFF`
2. `/Users/makinja/system/specs/john-drift-prevention-spec.md`
 - Status inside document: `CURRENT_RECONSTRUCTED_SPEC_PENDING_CEO_SIGNOFF`

What is already persisted

- Git repo: `/Users/makinja/system`
- Spec commit: `1a7b94a88 docs: reconstruct John fail-closed specs`
- Persistent evidence directory: `/Users/makinja/system/evidence/john-fail-closed-20260523/`
- Hash manifest for active gate files: `/Users/makinja/system/evidence/john-fail-closed-20260523/john-fail-closed-file-manifest-20260523.json`

Evidence to review

- Fresh Claude Code claim-gate evidence: `/tmp/claude-code-fresh-claim-gate-final-20260523.md`
- Reconstructed spec evidence: `/tmp/john-reconstructed-specs-20260523.md`
- MC review note: `/tmp/john-mc-review-notes-10599-10570-20260523.md`

- Memory memo: `/Users/makinja/.claude/projects/-Users-makinja/memory/project_john_fail_closed_claim_gates_2026-05-23.md`

MC state observed in this continuation

- MC #10599: `open`; follow-up evidence comment added; not marked ready/done.
- MC #10570: `paused`; follow-up evidence comment added; not marked ready/done.

BookStack sync status

- Attempted: `node /Users/makinja/system/tools/bookstack-sync.js sync`
- Result in this harness: PARTIAL / timed out after 120s.
- Do not treat BookStack as fully synced from this run.

Explicit non-claims

- Blueprint MUST workflow has not been started.
- MC #10599 and MC #10570 are not complete.
- CEO sign-off is not recorded yet.
- BookStack full sync is not confirmed.

If CEO signs off

Recommended next action after sign-off:

1. Record explicit sign-off in MC #10599 and #10570.
2. Decide whether to transition MC #10599 from `open` and MC #10570 from `paused` into normal workflow before `ready`.
3. Only run blueprint MUST wrapper/checklist if cost/risk approval is explicit.

Cost note

Last local cost probe in this continuation reported today: `$52.7932` across `58` Claude CLI requests.

John Blueprint Cost-Waiver Preflight — 2026-05-23

Blueprint MUST checklist — John fail-closed — 2026-05-23

Trigger / waiver

- CEO/user message: `Nije mi vazan cost.`
- Interpreted narrowly as: cost objection waived for continuing blueprint preflight/MUST work.
- Not interpreted as: CEO sign-off on specs, MC done approval, or permission to fabricate workflow state.

Inputs

- Project path: `/Users/makinja/system`
- MC #10599: John replacement plan
- MC #10570: John drift-prevention spec
- System BUILD-BLUEPRINT: `/Users/makinja/system/BUILD-BLUEPRINT.md`
- Persistent evidence base: `/Users/makinja/system/evidence/john-fail-closed-20260523/`

Deterministic checks run

1. Blueprint rubric wrapper:

- Command: `bash /Users/makinja/system/tools/run-blueprint-check.sh /Users/makinja/system --json`
- Evidence: `blueprint-check-rubric.json`
- Result: `READY`, score `100`, rc `0`.

2. Blueprint inventory wrapper:

- Command: `bash /Users/makinja/system/tools/run-blueprint-check.sh /Users/makinja/system --inventory --json`

- Evidence: `blueprint-check-inventory.json`
 - Result: `MISSING_FILES`, compliance `3/6`, rc `1`.
 - Missing: `DEPLOY-MAP.md`, `PIPELINE.md`, `RUNBOOK.md`.
 - Context: `/Users/makinja/system/BUILD-BLUEPRINT.md` says `~/system/` is exempt from `DEPLOY-MAP.md` and `RUNBOOK.md` circuit breakers because it is not a product. The generic inventory checker is not system-exemption-aware.
3. Blueprint SHA registry verify:
- Command: `node /Users/makinja/system/tools/blueprint-sha256-registry.js verify /Users/makinja/system/BUILD-BLUEPRINT.md`
 - Evidence: `blueprint-sha256-verify.out`, `blueprint-sha256-verify.err`
 - Result: `NOT_REGISTERED`, rc `2`.
 - Context: `/Users/makinja/.claude/agents/mehanik.md` CB9 defines `NOT_REGISTERED` / exit `2` as WARN, not block.
4. MC lease/status checks:
- Commands:
 - `node /Users/makinja/system/tools/mc.js claim-status 10599`
 - `node /Users/makinja/system/tools/mc.js claim-status 10570`
 - `node /Users/makinja/system/tools/mc.js show 10599`
 - `node /Users/makinja/system/tools/mc.js show 10570`
 - Evidence files: `mc-10599-claim-status.json`, `mc-10570-claim-status.json`, `mc-10599-show.txt`, `mc-10570-show.txt`
 - Result: both claim-status outputs show `available: true`, `lease_holder: null`.
5. Discover lookup:
- Command: `node /Users/makinja/system/tools/discover.js "john fail-closed claim gates system specs"`
 - Evidence: `discover-john-fail-closed.txt`
 - Result: lookup ran; output is advisory/context, not completion evidence.
6. Attempted Claude Code `/mehanik` skill wrapper via local cc-api-server:
- Tool: `run_claude_skill(skill="mehanik", ...)`
 - Result: BLOCKED by `cc-api-server error 401: Unauthorized`.
 - No mehanik marker was written from this attempt.

Verdict

PARTIAL / PRE-FLIGHT ONLY.

- Cost gate is waived by CEO/user statement.
- Deterministic local blueprint rubric passes.
- Inventory check fails generically but conflicts with documented `~/system/` exemption for non-product workspaces.
- SHA registry is WARN (`NOT_REGISTERED`), not block, under Mehanik CB9.
- Full Claude Code `/mehanik` wrapper did not run due cc-api-server 401.
- Therefore, do not claim full blueprint MUST completion.

Next required decision

Need explicit CEO decision on one of:

1. `SIGN OFF SPECS` — record sign-off, then move MC states via normal workflow.
2. `RUN FULL MEHANI/BLUEPRINT DESPITE 401` — use another authorized Claude Code path or manual operator run to generate `/tmp/mehanik-cleared-*` markers.
3. `REGISTER SYSTEM BUILD-BLUEPRINT HASH` — register `/Users/makinja/system/BUILD-BLUEPRINT.md` in SHA registry, then rerun preflight.
4. `REQUEST CHANGES` — edit specs/evidence before any status transition.