Pre-deal assessment: Project Larch

How the team behind Project Larch controls its work with coding agents, graded with the Prosights Agentic Engineering Index (AEI). The target ran AEI Scan on its own machines, gave its own answers and uploaded 3 results for 3 repositories. This site computed every score again from the check results. Each figure below says how well it is evidenced.

For the investment committee

Result	Overall score 82.5 to 86.4, median 85.9, of 100, over 3 results
Level	1, Written
Evidence	Weakest grade of any graded check: Declared. Share of the score at each grade: Declared 38.5%, Scanned 36.8%, Witnessed 22.8%, Verified by assessor 2%
Coverage	160 of 291 applicable checks are graded (55%)
Scope, as the target declared it	The target states 9 repositories with coding-agent use and uploaded results for 3. Nobody checked this statement.
Claim gap	5 checks where the target said yes and the scan found nothing
Open checks by severity	Critical 2, High 3, Medium 2, Low 11
Dates	Requested 2026-11-14; agreed 2026-11-14; computed 2026-11-14; the request ends 2026-12-31

In scope, in the target's words: "The 3 repositories of the payments product: payments-api, payments-web and ledger-service." Out of scope, in the target's words: "6 repositories of internal tools. 2 of them had coding-agent work this year."

The 5 most serious findings

Severity	Check	The practice the check asks for	Result	Evidence grade
Critical	9.1, level 3	The agent's sandbox is on.	open in 2 of 3 results: fail, claimed	Declared
Critical	9.2, level 3	The agent's network access is limited to a list.	open in 3 of 3 results: claimed, fail	Declared
High	1.4, level 3	A hook runs when a prompt is submitted to the agent.	open in 2 of 3 results: fail, claimed	Declared
High	6.3, level 3	A hook runs when the agent ends its turn.	open in 1 of 3 results: fail	Witnessed
High	15.11, level 3	No command can run outside the sandbox.	open in 3 of 3 results: claimed, fail	Declared

The 16 dimensions

#	Dimension	Score	Lowest level	Weakest grade
1	Requirements and recording	63 to 100, median 63	1, Written	Declared
2	Decisions	50 to 100, median 100	1, Written	Declared
3	Expected behavior and design	not rated	not rated	none
4	Agent briefing and model policy	70 to 100, median 70	1, Written	Declared
5	Build	70 to 100, median 100	1, Written	Declared
6	Test and evidence	50 to 100, median 100	2, Checked	Scanned
7	Review and merge	not rated	not rated	Declared
8	Release	not rated	not rated	Declared
9	Safety controls	63 to 81, median 81	1, Written	Declared
10	Memory and continuity	100	1, Written	Declared
11	Measurement	not rated	not rated	Declared
12	Governance	100	3, Enforced	Declared
13	Token and cost	64 to 100, median 81	1, Written	Declared
14	Harness depth	63 to 100, median 81	1, Written	Declared
15	Process guards	not rated	not rated	Declared
16	Model conduct	not rated	not rated	Declared

Where the results differ, a score reads "lowest to highest, median". 3 results from 3 repositories: no result is an average of the others.

Findings by severity

A finding is a check that is open: it failed, the target declined it, or the target claimed it with no evidence. The rule for each severity is written here, and it is the whole rule:

- Critical: A check of dimension 9, Safety controls, at level 1, 2 or 3 that failed or that the target declined to answer.
- High: A check where the target said yes and the scan found nothing; or a check at level 1 of any other dimension that failed or was declined.
- Medium: A check at level 2 outside dimension 9 that failed or was declined; or any other open check of dimension 9 (a level 4 check, or a claim with no evidence).
- Low: Every other open check: a failed or declined check at level 3 or 4, and a claim with no evidence, outside dimension 9.

Severity	Check	The practice the check asks for	What the check reads	Result and reason	Evidence grade
Critical	9.1, level 3	The agent's sandbox is on.	The sandbox setting	payments-api (Claude Code): fail; the setting is off / payments-web (Cursor): claimed; no documented location on Cursor; the answer was yes	Declared

Severity	Check	The practice the check asks for	What the check reads	Result and reason	Evidence grade
Critical	9.2, level 3	The agent's network access is limited to a list.	The setting that limits network access	ledger-service (Codex CLI): claimed; no documented location on Codex CLI; the answer was yes / payments-api (Claude Code): fail; the setting is not set / payments-web (Cursor): claimed; no documented location on Cursor; the answer was yes	Declared
High	1.4, level 3	A hook runs when a prompt is submitted to the agent.	A hook on the prompt-submit event (Claude Code and Codex CLI 'UserPromptSubmit'; VS Code 'userPromptSubmitted'; on other platforms, shown to a person)	payments-api (Claude Code): fail; no hook registered on UserPromptSubmit / payments-web (Cursor): claimed; no documented location on Cursor: no hook event for this is documented; the answer was yes	Declared
High	6.3, level 3	A hook runs when the agent ends its turn.	A hook on the end of the agent's turn (Claude Code and Codex CLI 'Stop'; Cursor 'stop'; VS Code 'agentStop'; on other platforms, shown to a person)	payments-api (Claude Code): fail; no hook registered on Stop	Witnessed
High	15.11, level 3	No command can run outside the sandbox.	The setting that lets a command run outside the sandbox	ledger-service (Codex CLI): claimed; no documented location on Codex CLI; the answer was yes / payments-api (Claude Code): fail; the setting is not set / payments-web (Cursor): claimed; no documented location on Cursor; the answer was yes	Declared
Medium	2.2, level 2	A check reads the decision records at commit.	The check that reads decision records at commit	payments-api (Claude Code): fail; assessor, 2026-11-12: Asked for; the team had no artifact to show.	Verified by assessor
Medium	9.5, level 3	When a hook itself fails, the action is refused.	What happens when the hook itself fails	ledger-service (Codex CLI): claimed; no documented location on Codex CLI: no setting for a failing hook is documented; the answer was yes	Declared
Low	16.4, level 2	A hook runs when a tool call is denied.	A hook on a denied tool call (Claude Code 'PermissionDenied'; Cursor 'postToolUseFailure'; on other platforms, shown to a person)	ledger-service (Codex CLI): claimed; no documented location on Codex CLI: no hook event for this is documented; the answer was yes	Declared

Severity	Check	The practice the check asks for	What the check reads	Result and reason	Evidence grade
Low	4.5, level 3	The agent's model is named with a version number.	The model named in settings and in agent definitions	ledger-service (Codex CLI): claimed; no documented location on Codex CLI; the answer was yes / payments-web (Cursor): claimed; no documented location on Cursor; the answer was yes	Declared
Low	4.6, level 3	A hook runs when an agent starts another agent.	A hook on the tool that starts another agent (Claude Code 'PreToolUse' on Agent; on other platforms, shown to a person)	ledger-service (Codex CLI): claimed; no documented location on Codex CLI: no hook event for this is documented; the answer was yes / payments-web (Cursor): claimed; no documented location on Cursor: no hook event for this is documented; the answer was yes	Declared
Low	5.2, level 3	A hook runs before the agent's shell commands.	A hook before shell commands (Claude Code 'PreToolUse' on Bash; Cursor 'beforeShellExecution'; on other platforms, shown to a person)	ledger-service (Codex CLI): claimed; no documented location on Codex CLI: no hook event for this is documented; the answer was yes	Declared
Low	13.3, level 3	A threshold is set at which the conversation is compacted.	The setting for compacting the conversation	payments-web (Cursor): claimed; no documented location on Cursor; the answer was yes	Declared
Low	13.4, level 3	The reasoning effort is set in settings or in an agent definition.	The reasoning-effort setting	payments-web (Cursor): claimed; no documented location on Cursor; the answer was yes	Declared
Low	13.6, level 3	A prompt cache lifetime is set, and caching is on.	The prompt-cache settings	ledger-service (Codex CLI): claimed; no documented location on Codex CLI; the answer was yes / payments-web (Cursor): claimed; no documented location on Cursor; the answer was yes	Declared
Low	14.1, level 3	A hook runs before the agent writes or edits a file.	A hook that runs before the agent writes or edits a file (Claude Code 'PreToolUse' on Write or Edit)	ledger-service (Codex CLI): claimed; no documented location on Codex CLI: no hook event for this is documented; the answer was yes / payments-web (Cursor): claimed; no documented location on Cursor: no hook event for this is documented; the answer was yes	Declared

Severity	Check	The practice the check asks for	What the check reads	Result and reason	Evidence grade
Low	14.3, level 3	An agent definition limits the tools its agent may use.	The tool limits in agent definitions	ledger-service (Codex CLI): claimed; no documented location on Codex CLI; the answer was yes	Declared
Low	11.1, level 4	Session records of the last 30 days are kept.	The count and dates of the agent's session records, never their text, and only when you turn the count on	payments-web (Cursor): claimed; no documented location on Cursor; the answer was yes	Declared
Low	13.12, level 4	Usage telemetry is on.	The usage-telemetry setting	ledger-service (Codex CLI): claimed; no documented location on Codex CLI; the answer was yes / payments-web (Cursor): claimed; no documented location on Cursor; the answer was yes	Declared

Questions for management

1 question for each check that has no result, that the target declined, or that rests on the target's own statement: 81 questions. Each asks to be shown what would pass the check.

Check	Why it is asked	Ask to be shown
1.1, level 1	rests on the target's own statement	Work items are written down before an agent starts on them. The check reads: Your answer to the survey question "Are work items written down before an agent starts on them?".
1.2, level 2	not answered or not readable	A check refuses a change that is tied to no work item. The check reads: The check that refuses a change tied to no item.
1.3, level 2	not answered or not readable	A script regenerates the list of open work. The check reads: A list of open work that a script regenerates.
1.4, level 3	rests on the target's own statement	A hook runs when a prompt is submitted to the agent. The check reads: A hook on the prompt-submit event (Claude Code and Codex CLI 'UserPromptSubmit'; VS Code 'userPromptSubmitted'; on other platforms, shown to a person).
2.2, level 2	not answered or not readable	A check reads the decision records at commit. The check reads: The check that reads decision records at commit.
2.3, level 1	rests on the target's own statement	It is written which choices a person makes and which an agent may make. The check reads: Your answer to the survey question "Is it written which choices a person makes and which an agent may make?".
3.1, level 1	not answered or not readable	The expected behaviors of a change are written before the build. The check reads: For 1 recent change: the expected behaviors written before the build.
3.2, level 2	not answered or not readable	A check ties each expected behavior to a test. The check reads: The check that ties each expected behavior to a test.

Check	Why it is asked	Ask to be shown
4.3, level 1	rests on the target's own statement	Rule files apply only to some paths. The check reads: Rule files that apply only to some paths.
4.5, level 3	rests on the target's own statement	The agent's model is named with a version number. The check reads: The model named in settings and in agent definitions.
4.6, level 3	rests on the target's own statement	A hook runs when an agent starts another agent. The check reads: A hook on the tool that starts another agent (Claude Code 'PreToolUse' on Agent; on other platforms, shown to a person).
5.1, level 1	rests on the target's own statement	Each agent works in its own checkout or environment. The check reads: Your answer to the survey question "Does each agent work in its own checkout or environment?".
5.2, level 3	rests on the target's own statement	A hook runs before the agent's shell commands. The check reads: A hook before shell commands (Claude Code 'PreToolUse' on Bash; Cursor 'beforeShellExecution'; on other platforms, shown to a person).
5.3, level 2	not answered or not readable	A limit stops a build or test run that hangs. The check reads: The limit that stops a build or test run that hangs.
5.4, level 1	rests on the target's own statement	The size of 1 agent task has a written limit. The check reads: Your answer to the survey question "Is there a written limit on the size of 1 agent task?".
6.4, level 4	not answered or not readable	The test results of each build are kept as files. The check reads: Test results kept as files for each build.
6.5, level 4	not answered or not readable	A test set for each crucial function runs at every release. The check reads: For each function you name as crucial: a test set run at every release.
6.6, level 2	not answered or not readable	A gate fails on new warnings in changed files. The check reads: The gate that fails on new warnings in changed files.
7.1, level 1	rests on the target's own statement	A reviewer other than its author reviews every agent change before it merges: a person, or a model run as a separate reviewer. The check reads: Your answer to the survey question "Is every agent change reviewed before it merges by a reviewer other than its author: a person, or a model run as a separate reviewer?".
7.3, level 2	not answered or not readable	An automated review runs on every change before it merges and cannot be skipped. The check reads: An automated review that runs on every change before it merges and cannot be skipped.
7.4, level 2	not answered or not readable	A merge step builds and checks after each merge and stops on failure. The check reads: The merge step that builds and checks after each merge and stops on failure.
7.5, level 1	rests on the target's own statement	It is written who may start or assign an agent. The check reads: Your answer to the survey question "Is it written who may start or assign an agent?".

Check	Why it is asked	Ask to be shown
7.6, level 3	not answered or not readable	A written rule says which changes need a person's approval before they merge or ship, by risk. The check reads: The written rule that says which changes need a person's approval before they merge or ship, by risk.
7.7, level 4	not answered or not readable	Automated approvals are recorded, and a person reviews a sample of them each quarter. The check reads: The record of automated approvals, and the sample of them a person reviewed in the last quarter.
8.1, level 1	rests on the target's own statement	A person confirms each release. The check reads: Your answer to the survey question "Does a person confirm each release?".
8.2, level 2	not answered or not readable	A release gate refuses when a required item is missing. The check reads: The release gate that refuses when a required item is missing.
8.3, level 4	not answered or not readable	The gate result of each release is recorded. The check reads: The record of each release's gate result.
9.1, level 3	rests on the target's own statement	The agent's sandbox is on. The check reads: The sandbox setting.
9.2, level 3	rests on the target's own statement	The agent's network access is limited to a list. The check reads: The setting that limits network access.
9.5, level 3	rests on the target's own statement	When a hook itself fails, the action is refused. The check reads: What happens when the hook itself fails.
9.6, level 2	not answered or not readable	A secret scan runs before code leaves the machine. The check reads: The secret scan that runs before code leaves the machine.
9.8, level 4	not answered or not readable	A list marks each of the team's rules as enforced, checked or written only. The check reads: A list of the team's rules marked enforced, checked or written only.
10.1, level 1	rests on the target's own statement	The agent keeps memory files for this project. The check reads: How many files the agent's memory folder holds, never their contents.
10.2, level 2	not answered or not readable	An end-of-session step refuses to finish with failing tests or unpushed work. The check reads: The end-of-session step that refuses to finish with failing tests or unpushed work.
10.3, level 1	rests on the target's own statement	The start-of-session procedure is written. The check reads: Your answer to the survey question "Is the start-of-session procedure written?".
10.4, level 1	rests on the target's own statement	Each agent writes down what it learned from its work, in a place the next agent reads. The check reads: Your answer to the survey question "Does each agent write down what it learned from its work, in a place the next agent reads?".
10.5, level 2	not answered or not readable	A mechanism refuses a change whose agent recorded no lesson entry. The check reads: The mechanism that refuses a change whose agent recorded no lesson entry.

Check	Why it is asked	Ask to be shown
10.6, level 2	not answered or not readable	For each of the last 5 agent changes, a record shows that the agent read the carried-forward lessons before its work. The check reads: For the last 5 agent changes, the record that the agent read the carried-forward lessons before its work.
11.1, level 4	rests on the target's own statement	Session records of the last 30 days are kept. The check reads: The count and dates of the agent's session records, never their text, and only when you turn the count on.
11.2, level 4	not answered or not readable	The merge rate and the revert rate of agent changes are measured against the team's own baseline. The check reads: Merge rate and revert rate of agent changes against the team's own baseline.
11.3, level 4	not answered or not readable	Usage or cost is recorded per agent or task. The check reads: Usage or cost recorded per agent or task.
11.4, level 4	not answered or not readable	A quality trend is kept over time: warnings, complexity or failures. The check reads: A quality trend: warnings, complexity or failures over time.
11.5, level 1	rests on the target's own statement	Agent work is measured by a named outcome. The check reads: Your answer to the survey question "Do you measure agent work by a named outcome?".
12.1, level 1	rests on the target's own statement	A written statement says which AI tools are used, on which data, and who approves. The check reads: Your answer to the survey question "Is there a written statement of which AI tools are used, on which data, and who approves?".
12.3, level 1	not answered or not readable	The rules have stable identifiers and a record of why each exists. The check reads: Rules with stable identifiers and a record of why each exists.
12.4, level 1	rests on the target's own statement	Agent use is covered by the organization's AI risk process. The check reads: Your answer to the survey question "Is agent use covered by your organization's AI risk process?".
13.1, level 1	rests on the target's own statement	It is written which model and which reasoning effort each kind of task uses. The check reads: Your answer to the survey question "Is it written which model and which reasoning effort each kind of task uses?".
13.2, level 1	not answered or not readable	Every instruction file has fewer than 200 lines. The check reads: The number of lines in each instruction file, never the lines.
13.3, level 3	rests on the target's own statement	A threshold is set at which the conversation is compacted. The check reads: The setting for compacting the conversation.
13.4, level 3	rests on the target's own statement	The reasoning effort is set in settings or in an agent definition. The check reads: The reasoning-effort setting.
13.6, level 3	rests on the target's own statement	A prompt cache lifetime is set, and caching is on. The check reads: The prompt-cache settings.

Check	Why it is asked	Ask to be shown
13.7, level 4	not answered or not readable	A report separates input tokens read from the cache from input tokens sent uncached. The check reads: A report that separates input tokens read from the cache from input tokens sent uncached.
13.8, level 2	not answered or not readable	A mechanism shortens long command output before the agent reads it. The check reads: A mechanism that shortens long command output before the agent reads it.
13.9, level 1	rests on the target's own statement	It is written that long or verbose work goes to its own agent, and that agents wait without polling. The check reads: Your answer to the survey question "Is it written that long or verbose work goes to its own agent, and that agents wait without polling?".
13.10, level 3	not answered or not readable	A spend or usage limit is set at the provider. The check reads: A spend or usage limit set at the provider, for the workspace or for each person.
13.11, level 4	not answered or not readable	Tokens or cost per model are reviewed over time, and a change was made because of a review. The check reads: A review of tokens or cost per model over time, and 1 change made because of it.
13.12, level 4	rests on the target's own statement	Usage telemetry is on. The check reads: The usage-telemetry setting.
14.1, level 3	rests on the target's own statement	A hook runs before the agent writes or edits a file. The check reads: A hook that runs before the agent writes or edits a file (Claude Code 'PreToolUse' on Write or Edit).
14.3, level 3	rests on the target's own statement	An agent definition limits the tools its agent may use. The check reads: The tool limits in agent definitions.
14.5, level 2	not answered or not readable	A command shows the state of each background or unattended run. The check reads: The command that shows the state of each background or unattended run.
14.6, level 3	not answered or not readable	An agent runs commands on another machine only from an allow list. The check reads: The channel through which an agent runs a command on another machine.
14.7, level 2	not answered or not readable	The team's own hooks and guard scripts have tests. The check reads: The tests of the team's own hooks and guard scripts.
15.1, level 3	not answered or not readable	Every delete goes through 1 script, and a hook refuses a delete made any other way. The check reads: The script every delete goes through, and the hook that refuses a delete made any other way.
15.2, level 3	not answered or not readable	A hook refuses a build or test command run outside the team's wrapper script. The check reads: The hook that refuses a build or test command run outside the team's wrapper script.
15.3, level 2	not answered or not readable	A check keeps credentials and real messages out of test fixtures. The check reads: The check that keeps credentials and real messages out of test fixtures.

Check	Why it is asked	Ask to be shown
15.4, level 1	rests on the target's own statement	It is written that a new integration or tool starts with no access until it is listed. The check reads: Your answer to the survey question "Is it written that a new integration or tool starts with no access until it is listed?".
15.5, level 2	not answered or not readable	A mechanism stops a test from reaching a live account unless a person turns that on. The check reads: What stops a test from reaching a live account unless a person turns that on.
15.6, level 2	not answered or not readable	A commit check refuses personal identifiers and authorship marks. The check reads: The commit check that refuses personal identifiers and authorship marks.
15.7, level 2	not answered or not readable	A set of injection attempts is run against a model before it is given a channel. The check reads: The set of injection attempts run against a model before it is given a channel, and the step that runs it.
15.8, level 1	rests on the target's own statement	It is written that a change touches only what its own work item needs. The check reads: Your answer to the survey question "Is it written that a change touches only what its own work item needs?".
15.9, level 2	not answered or not readable	A guard has a single switch, and its off state is tested. The check reads: For 1 guard: its single switch, and the test of the off state.
15.10, level 3	not answered or not readable	A script removes an agent's workspace and refuses when its work is not merged. The check reads: The script that removes an agent's workspace and refuses when its work is not merged.
15.11, level 3	rests on the target's own statement	No command can run outside the sandbox. The check reads: The setting that lets a command run outside the sandbox.
16.1, level 1	rests on the target's own statement	A written record shows how each model in use has failed in the team's own work. The check reads: Your answer to the survey question "Is there a written record of how each model you use has failed in your own work?".
16.2, level 1	rests on the target's own statement	Every brief to an agent states what done means. The check reads: Your answer to the survey question "Does every brief to an agent state what done means?".
16.3, level 2	not answered or not readable	Someone other than the agent checks the claims in its reports. The check reads: For the last 2 agent reports: who other than the agent checked the claims, and what they checked.
16.4, level 2	rests on the target's own statement	A hook runs when a tool call is denied. The check reads: A hook on a denied tool call (Claude Code 'PermissionDenied'; Cursor 'postToolUseFailure'; on other platforms, shown to a person).
16.5, level 2	not answered or not readable	Refused calls are logged, each with the model that made the call. The check reads: The log of refused calls, each with the model that made the call.

Check	Why it is asked	Ask to be shown
16.6, level 3	not answered or not readable	A mechanism refuses the end of a turn while tasks are open. The check reads: The mechanism that refuses the end of a turn while tasks are open, and 1 refusal it made.
16.8, level 4	not answered or not readable	Rule breaks, early stops and false "done" reports are counted per model, and the counts are read on a schedule. The check reads: Counts per model for 2 periods (rule breaks, early stops, false "done" reports) and the schedule on which they are read.
16.9, level 4	not answered or not readable	A new model version was measured before it was adopted. The check reads: The measurement made before the last new model version was adopted, and the line of the model policy it changed or confirmed.

Remediation plan

Every open check, the lowest level first, and within a level by the kind of work. The 3 kinds of work:

- Write it down. The check passes on a written answer: the team writes the rule or names the place.
- Add a setting or a file. The check passes when a setting or a file exists at a location the coding agent's vendor documents.
- Build a mechanism and show it. The check passes when a person is shown a working check, gate, script or record.

No estimate of hours or cost is given: none was measured. The target's own report from AEI Scan names the exact setting for each scan check.

Level	Kind of work	Check	What to put in place	Where	Open in
2	Build a mechanism and show it	2.2	Add a check that reads the decision records at commit.	shown to a person	1 of 3
2	Build a mechanism and show it	16.4	Register a hook that runs when a tool call is denied.	shown to a person	1 of 3
3	Add a setting or a file	9.1	Turn the agent's sandbox on.	.claude/settings.json, .claude/settings.local.json, ~/.claude/settings.json	2 of 3
3	Add a setting or a file	9.2	Limit the agent's network access to a list.	.claude/settings.json, .claude/settings.local.json, ~/.claude/settings.json	3 of 3
3	Add a setting or a file	1.4	Register a hook that runs when a prompt is submitted to the agent.	.claude/settings.json, .claude/settings.local.json, ~/.claude/settings.json	2 of 3
3	Add a setting or a file	6.3	Register a hook that runs a check when the agent ends its turn.	.claude/settings.json, .claude/settings.local.json, ~/.claude/settings.json	1 of 3
3	Add a setting or a file	15.11	Set the sandbox so that no command can run outside it.	.claude/settings.json, .claude/settings.local.json, ~/.claude/settings.json	3 of 3
3	Build a mechanism and show it	9.5	Set every hook so that the action is refused when the hook itself fails.	shown to a person	1 of 3
3	Build a mechanism and show it	4.5	Name the agent's model with a version number, not by an alias.	shown to a person	2 of 3

Level	Kind of work	Check	What to put in place	Where	Open in
3	Build a mechanism and show it	4.6	Register a hook on the tool that starts another agent.	shown to a person	2 of 3
3	Build a mechanism and show it	5.2	Register a hook that runs before the agent's shell commands.	shown to a person	1 of 3
3	Build a mechanism and show it	13.3	Set a threshold at which the agent's conversation is compacted.	shown to a person	1 of 3
3	Build a mechanism and show it	13.4	Set the agent's reasoning effort in settings or in an agent definition.	shown to a person	1 of 3
3	Build a mechanism and show it	13.6	Set a prompt cache lifetime, and leave caching on.	shown to a person	2 of 3
3	Build a mechanism and show it	14.1	Register a hook that runs before the agent writes or edits a file.	shown to a person	2 of 3
3	Build a mechanism and show it	14.3	Limit the tools of at least 1 agent in its agent definition.	shown to a person	1 of 3
4	Build a mechanism and show it	11.1	Keep the agent's session records for at least 30 days.	shown to a person	1 of 3
4	Build a mechanism and show it	13.12	Turn usage telemetry on for the agent.	shown to a person	2 of 3

How well each figure is evidenced

Grade	What it means	Graded checks	Share of the score
Declared	The target said so: a survey answer, or the target's own statement that an artifact exists. Nobody outside the target saw it.	79	38.5%
Scanned	AEI Scan read it on the target's machine, in a run that nobody outside the target watched.	43	36.8%
Witnessed	AEI Scan read it in a run that an assessor watched. The assessor recorded the date and what was watched.	35	22.8%
Verified by assessor	An assessor was shown the artifact and graded the check. The assessor recorded the date and what was shown.	3	2%

A check that failed carries no points, so a grade can hold many checks and a small share. A witnessed run shows the repository at the moment of the run, and a team that is watched may behave differently. An unobserved run can be staged.

The results, 1 for each repository and coding agent

Repository	Coding agent	Score	Level	Coverage	Claim gap	Weakest grade	Run
ledger-service	Codex CLI	86.4	1, Written	54%	0	Declared	Not watched
payments-api	Claude Code	82.5	1, Written	58%	5	Declared	Witnessed 2026-11-12: A screen share of the whole run, started from a fresh clone of the repository.
payments-web	Cursor	85.9	1, Written	54%	0	Declared	Not watched

Appendix: integrity of each file

Repository	Coding agent	SHA-256 of the uploaded file	Tool	Catalog	Run (UTC)	Uploaded (UTC)
ledger-service	Codex CLI	e595e1595b89d11b6253cad3bbb10be79b69a99a785c7746960a7bb9773f3d4d	AEI Scan 0.1.0	2026-11-16	2026-11-16 19:42	2026-11-14 12:00
payments-api	Claude Code	056178bc7584389c92c47a8249787bec1787a6c7c67f897a90644dc03ec59fb2	AEI Scan 0.1.0	2026-11-16	2026-11-16 19:42	2026-11-14 12:00
payments-web	Cursor	28eef9f2aa1fd0fe9e7e42929142625fd100add4409ef6b7a3db11ab23957ac8	AEI Scan 0.1.0	2026-11-16	2026-11-16 19:42	2026-11-14 12:00

The SHA-256 shows that the stored file is the file AEI Scan printed that value for. It does not show who ran the scan or on which machine: the file carries no signature, and the results file does not record the operating system. The names of the repositories are the folder names the uploads gave.

Appendix: said but not found, and in place but not known

Said but not found: the target answered yes and the scan found nothing.

- ledger-service (Codex CLI): none
- payments-api (Claude Code): 1.4, 6.3, 9.1, 9.2, 15.11
- payments-web (Cursor): none

In place but not known: the scan found it and the target answered no.

- ledger-service (Codex CLI): none
- payments-api (Claude Code): none
- payments-web (Cursor): none

Appendix: token and cost (dimension 13)

Check	The practice the check asks for	ledger-service, Codex CLI	payments-api, Claude Code	payments-web, Cursor
13.1, level 1	It is written which model and which reasoning effort each kind of task uses.	pass (Declared)	pass (Declared)	pass (Declared)
13.2, level 1	Every instruction file has fewer than 200 lines.	unknown	pass (Witnessed)	unknown
13.3, level 3	A threshold is set at which the conversation is compacted.	pass (Scanned)	pass (Witnessed)	claimed (Declared)

Check	The practice the check asks for	ledger-service, Codex CLI	payments-api, Claude Code	payments-web, Cursor
13.4, level 3	The reasoning effort is set in settings or in an agent definition.	pass (Scanned)	pass (Witnessed)	claimed (Declared)
13.5, level 3	Agents that another agent starts are given their own model.	pass (Scanned)	pass (Witnessed)	pass (Scanned)
13.6, level 3	A prompt cache lifetime is set, and caching is on.	claimed (Declared)	pass (Witnessed)	claimed (Declared)
13.7, level 4	A report separates input tokens read from the cache from input tokens sent uncached.	unknown	unknown	unknown
13.8, level 2	A mechanism shortens long command output before the agent reads it.	unknown	unknown	unknown
13.9, level 1	It is written that long or verbose work goes to its own agent, and that agents wait without polling.	pass (Declared)	pass (Declared)	pass (Declared)
13.10, level 3	A spend or usage limit is set at the provider.	unknown	unknown	unknown
13.11, level 4	Tokens or cost per model are reviewed over time, and a change was made because of a review.	unknown	unknown	unknown
13.12, level 4	Usage telemetry is on.	claimed (Declared)	pass (Witnessed)	claimed (Declared)

Appendix: model conduct (dimension 16)

Check	The practice the check asks for	ledger-service, Codex CLI	payments-api, Claude Code	payments-web, Cursor
16.1, level 1	A written record shows how each model in use has failed in the team's own work.	pass (Declared)	pass (Declared)	pass (Declared)
16.2, level 1	Every brief to an agent states what done means.	pass (Declared)	pass (Declared)	pass (Declared)
16.3, level 2	Someone other than the agent checks the claims in its reports.	unknown	unknown	unknown
16.4, level 2	A hook runs when a tool call is denied.	claimed (Declared)	pass (Witnessed)	pass (Scanned)
16.5, level 2	Refused calls are logged, each with the model that made the call.	unknown	unknown	unknown
16.6, level 3	A mechanism refuses the end of a turn while tasks are open.	unknown	unknown	unknown
16.7, level 3	A hook runs when a started agent ends its work or a task is marked complete.	pass (Scanned)	pass (Witnessed)	pass (Scanned)
16.8, level 4	Rule breaks, early stops and false "done" reports are counted per model, and the counts are read on a schedule.	unknown	unknown	unknown
16.9, level 4	A new model version was measured before it was adopted.	unknown	unknown	unknown

The counts below are the target's own, for each model its agents run on: how often the model broke a rule, stopped early or asked for a permission it had, reported done when it was not, invented a reason, or did work nobody asked for. AEI Scan counted none of them and they change no grade. Their evidence grade is Declared.

Repository	Model	Period	Rules broken	Stopped early or asked for a permission it had	Reported done when it was not	Invented a reason	Did work nobody asked for
payments-api	Model A	2026-09	9	6	3	2	4
payments-api	Model A	2026-10	4	7	1	2	5
payments-api	Model B	2026-10	2	1	2		

Appendix: how the results compare

Repository	Coding agent	Score	A default installation of this coding agent	Among all shared results
ledger-service	Codex CLI	86.4	passes 0 of 35 scan checks; this result passes 23	no comparison: 0 shared results, fewer than 11
payments-api	Claude Code	82.5	passes 0 of 35 scan checks; this result passes 30	no comparison: 0 shared results, fewer than 11
payments-web	Cursor	85.9	passes 0 of 35 scan checks; this result passes 22	no comparison: 0 shared results, fewer than 11

The default installation is the coding agent as installed, with nothing configured (catalog of 2026-11-16). The shared results are those whose owners chose to share them; they are not a sample of all teams, and a figure is shown only for a group of at least 11.

What this report does not show

- No proven link to delivery speed or code quality. No study read for the framework relates an AEI score to either.
- Presence, not effect. A passed check shows that a control exists or was shown; it does not show how well the control works.
- Not a security audit, a code-quality review or a licensing review. AEI Scan reads settings, file names and hook registrations; it opens no source file.
- Not the whole company. The scope is the target's own statement and nobody checked it; a repository the target left out is not in this report.
- Not an observed run, unless a result says Witnessed. A run that nobody outside the target watched can be staged: a repository can be prepared for the scan.
- Not a statement about what the company is worth. This report makes no claim about value, returns or price.