

Evaluator Roster

The roster names who — human or automated — is allowed to sign off on what, when they are invoked, the scope of their authority, what artifacts their sign-off produces, and where escalation goes when they cannot or should not decide alone.

A sign-off is only valid when recorded in the evaluator_signatures array of an s2-audit-trail-schema provenance record. Unsigned promotion is a violation of the evaluation gate.

Roster Matrix

Role	When Invoked	Authority Scope	Sign-Off Requirement	Escalation Path
builder-self-test (human)	Before any PR to a GROW artifact; on every local change to prompts, tools, or workflow steps.	Functional family only; no n-high-stakes systems only.	One signature in evaluator_signatures with role=builder-self-test, verdict ∈ {pass, fail}. May NOT sign off on safety or quality families.	Any safety fail, any high-stakes system, or any artifact touching regulator-visible output escalates to domain-expert-reviewer and compliance-reviewer.
automated-rubric-harness (automated)	On every eval run, every regression run, and on a nightly cadence against the frozen reference.	Functional, quality, and edge-case families. Safety family in advisory mode only — automated verdicts here are inputs, never gates.	Machine signature with role=automated-rubric-harness, actor=harness@<version>, verdict, and a link to the scorecard fragment in s2-scoring-system.	Disagreement with human-evaluator on the same artifact triggers domain-expert-reviewer adjudication. Three consecutive harness failures on the same failure_id triggers incident path.
human-evaluator (human)	On every release candidate for a system operating at operational quality; on every artifact touching a high-stakes decision per s2-evaluation-charter.	Quality and safety families; binding sign-off for functional family for high-stakes systems.	Two-of-two when the system is high-stakes (one human-evaluator plus one domain-expert-reviewer); one signature otherwise. Verdict + free-text rationale required.	Confidence-band conflicts (e.g., agent cited high confidence on a claim the evaluator judges low) escalate to domain-expert-reviewer.

Role	When Invoked	Authority Scope	Sign-Off Requirement	Escalation Path
domain-expert-reviewer (human, named)	When the system operates inside a regulated or specialist domain — Florida special-district statutes, FASD compliance, STR tax (FL transient rental), DEP grant covenants, podcast IP/licensing, IRS treatment for the venture studio.	Binding sign-off on safety tests that depend on domain interpretation; veto power on any release that would emit a domain claim the expert cannot verify.	One signature with role=domain-expert-reviewer, named actor (not "the domain team"), scope field naming the statute, grant, or contract family.	Disagreement between two domain experts (e.g., FS § 720.3085 reading vs. § 718) escalates to external-auditor. Public-comment or regulator-bound output additionally routes through compliance-reviewer.
compliance-reviewer (human, named)	On every artifact that may carry PII, that constitutes a public record under FL Sunshine, that touches DEP/FDOT/HOA correspondence, that quotes a grant ID, or that handles guest financial data.	Binding sign-off on safety tests seeded from failure modes tagged pii, sunshine, grant, tax, or fair-housing. May override a builder pass with a fail.	One signature plus a retention-policy attachment (which evidence is kept, for how long, where).	A safety fail tagged critical or any suspected statute violation routes to external-auditor and triggers the incident path in s2-regression-discipline.
external-auditor (third party)	On scheduled cadence (quarterly for high-stakes systems; annually for the GROW library as a whole); on demand after a critical incident; before any public claim that a GROW system meets a named external standard (e.g., SOC 2 scope, NIST AI RMF alignment).	Read-only access to all provenance records in scope; advisory and reporting only — does not sign off on individual releases.	Issues a written report attached to the audit package per s2-audit-package-templates. Findings open regression-discipline change-log entries that the library owner must resolve or formally accept as risk.	Findings classified material escalate to the GROW library owner (Doug Liles) for written acceptance or remediation plan within 30 days.

Invocation Rules

1. No self-clearing on high-stakes. A builder may never be the sole signatory on an artifact destined for regulator-visible, financially binding, or irreversible output. The Charter overrides individual roster permissions.
2. Named actors only. `evaluator_signatures.actor` MUST resolve to a specific person (`evaluator:doug.liles`) or a specific harness version (`harness@0.2.1`). Generic role names ("the team") are rejected by the audit-trail validator.
3. Two-of-two for high-stakes. Human-evaluator + domain-expert-reviewer is the standing requirement for any artifact flagged high-stakes by the Charter.
4. Compliance has veto. A compliance-reviewer fail cannot be overridden by a builder or domain expert. It is overridden only by a written external-auditor finding or by formal risk acceptance signed by the library owner.

Authority Scope — Examples

- A builder testing a podcast chapter-titling agent at demo quality: builder-self-test alone is sufficient.
- The same agent, before it auto-publishes to Spotify: add automated-rubric-harness and human-evaluator; quality family becomes binding.
- A permit-triage agent draft response that cites FS § 161: add domain-expert-reviewer and compliance-reviewer; two-of-two required.
- A new vendor MSA review tool for the venture studio: domain-expert-reviewer (legal) plus compliance-reviewer; external-auditor on next quarterly cycle.

Escalation Paths — Defaults

- Critical safety fail → compliance-reviewer + library owner within 24 hours; regression-discipline entry opened same day.
- Confidence-band conflict → domain-expert-reviewer; resolution recorded in `decision_trace` with `decision_origin = human-override` when the human re-rules.
- Harness vs. human disagreement → domain-expert-reviewer arbitrates; the binding verdict is recorded; both inputs retained in the provenance record.
- External-auditor material finding → library owner; 30-day written response.

Maintenance

The roster is reviewed at every quarterly cadence and whenever a new high-stakes domain enters scope (e.g., a new grant program, a new STR jurisdiction, a new podcast distribution surface). Changes are logged via `s2-regression-discipline`.