

Interaction Conformance Standard

The readback layer, operationalized. Version 0.1, July 2026. Validation of agent-built interaction design as a formal part of design system governance: written to survive a change control board and a quarterly budget review. Adoptable as written. tokenstoagents.ai/standard

Scope. All UI and interaction design produced wholly or partly by AI agents for governed surfaces (web, native, kiosk, agent desktop, embedded), every locale shipped. Co-produced work is covered regardless of authorship split. **Definition.** The readback layer is the criteria, machinery, sign-offs, and failure handling that verify agent-built work against the design system before and after ship. It validates behavior, not just pixels.

Conformance criteria, pass or fail, in this order

F. Foundations	Approved tokens only; no hard-coded values; locale token layers respected (script floors, density, RTL).
C. Components	Real system components imported from system packages; no lookalikes; published APIs; deprecated blocked.
I. Interaction	Flows and states match sanctioned patterns: navigation, forms, errors, empty states, loading, motion. No invented patterns without an approved proposal. Machines check structure; humans judge.
A. Accessibility	WCAG 2.2 AA; EN 301 549 mapping for EU surfaces; acceptance criteria on the originating ticket.
L. Locale	Strings externalized with context; ICU valid; expansion tolerances respected; blocking locales before merge; no locale forks.

Gates

V1, automated, every change: token lint, component provenance, accessibility engine, visual regression incl. tall-script, long-word, and RTL screens, string and ICU validation. Failures bounce without human time. **V2, benchmark cadence:** the agent context (MCP, rules, orchestration files) is benchmarked monthly and on any change, tracking conformance and hallucination rates; context changes are API changes, human-approved. **V3, human sign-off, non-delegable:** every merge, deprecation, breaking change, exception grant, regulated locale sign-off, and rule change. Sign-off happens on the readback. **V4, audit sampling:** monthly human audit of shipped agent work against all five classes, including raw output behind clean readbacks.

The readback

Every run produces the readback: a checker agent, never the builder, on independent context, reads the application back against the system and renders a monochrome scaffold where the only color on the page is the color that shouldn't be there. Components carry system names; lookalikes flagged; flows and states mapped to the patterns they claim; open exceptions carried forward. Facts come from deterministic checks; the agent assembles and narrates, never adjudicates. The readback is evidence, not absolution: every red item resolves, and V4 measures the readback's own false-negative rate. One agent builds, a different agent reads it back, a human signs: maker-checker, applied to interaction design. The readback attaches to the change record.

Failure handling

Three outcomes, no fourth: block; fix (re-gated); or exception, entered in the register with an owner and an expiry, then promoted, migrated, or deprecated at review. Silent drift is a governance incident, not a design preference.

Roles and metrics

Roles: platform engineer owns V1, V2, the checker agent, and the readback pipeline; the core lead owns exceptions and V4; designers review readbacks (the designer's V3 interface); accessibility owns class A; locale owners own class L; the council owns this standard, class I pattern approvals, and the monthly register review. Auto-merge only for token-level fixes with passing tests on non-regulated surfaces, a council policy switch. **Quarterly metrics with dollar attribution:** conformance rate at first gate; foundation drift rate; readback false-negative rate; hallucination rate per benchmark; mean time to detect drift; exception count and age; coverage as a governed ratio (about 70 percent new-feature, 50 percent existing).

Cadence and maturity: V1 every change; V2 monthly; V3 continuous, on the readback; V4 monthly. Standard reviewed quarterly; re-baselined annually against accessibility law and the token specification. Adoption presumes published decision rights, contribution criteria, and a deprecation policy: enforcement without documented decision rights is theater.