Agent Logic

UTS v1.1 reference baseline
Preprint in review · July 2026

Universal Tool Schema (UTS)

A provider-neutral standard for describing what AI tools actually do.

THE PROBLEM

Tool calls are where agents touch the real world — and today's tool definitions stop at **name, description, parameters**. That is enough to call a function, but not enough to decide whether the call is safe to retry, safe to replay, safe to parallelize, or safe to allow at all. Every provider format leaves the operational facts buried in prose, prompts, or glue code. As agents reach production, that gap becomes the audit finding.

WHAT UTS DECLARES

- **Side effects** — does the tool read, write, mutate, delete, or send data outside the system?
- **Replay posture** — are retry, replay, batching, and parallel execution safe? Is output deterministic?
- **Requirements** — what authentication, resources, and execution environment must exist first?
- **Sensitivity & observability** — what data does it touch, and what should reviewers and logs see?

Per-call invocation metadata travels alongside; conformance fixtures — valid, invalid, and *dangerous-but-valid* — make schema behavior testable rather than aspirational.

THE DESIGN RULE

> ### UTS validity ≠ execution authority
>
> UTS describes; it never authorizes. A valid UTS object is a valid description — permission to execute stays with the runtime, policy engine, or governance layer. That boundary is what makes UTS adoptable: it standardizes the facts everyone needs without dictating anyone's authorization model.

EVIDENCE

15	14/15	11/11
MODELS TESTED	PERFECT BOTH LANES	TASKS, UTS-ONLY

On a fixed 11-task benchmark across **5 hosted models** (OpenAI, Google, Anthropic) and **10 open/local models**, 14 of 15 scored 11/11 in both lanes: ordinary provider-style tool calls *and* portable UTS-only calls. The exception was one 4B-class small model (8/11 UTS-only).

The practical finding: UTS costs essentially nothing in model capability. Current models emit portable UTS tool calls as reliably as provider-native formats.

Every published row links to promoted run artifacts (JSON results, run logs, provider status, self-check output), and a deterministic self-check keeps the benchmark reproducible. Benchmark date: 2026-05-21.

STATUS & BOUNDARY

UTS v1.1 is the implemented reference baseline: JSON Schemas, examples (including OpenAI-compatible and MCP-style wrappers), conformance fixtures, and indexed evidence. Apache-2.0. A preprint is in review.

Benchmark results establish format capability on the stated panel — they are not provider endorsements or production-safety claims.

WHY IT MATTERS

Whoever standardizes the safety vocabulary of tool calling owns a durable layer of agent infrastructure. UTS is positioned to be that layer: neutral enough for anyone to adopt, rigorous enough for enterprises to audit against, and already backed by reproducible evidence.

hello@agent-logic.ai