AGENT TASK EVALUATION

TRACE Framework

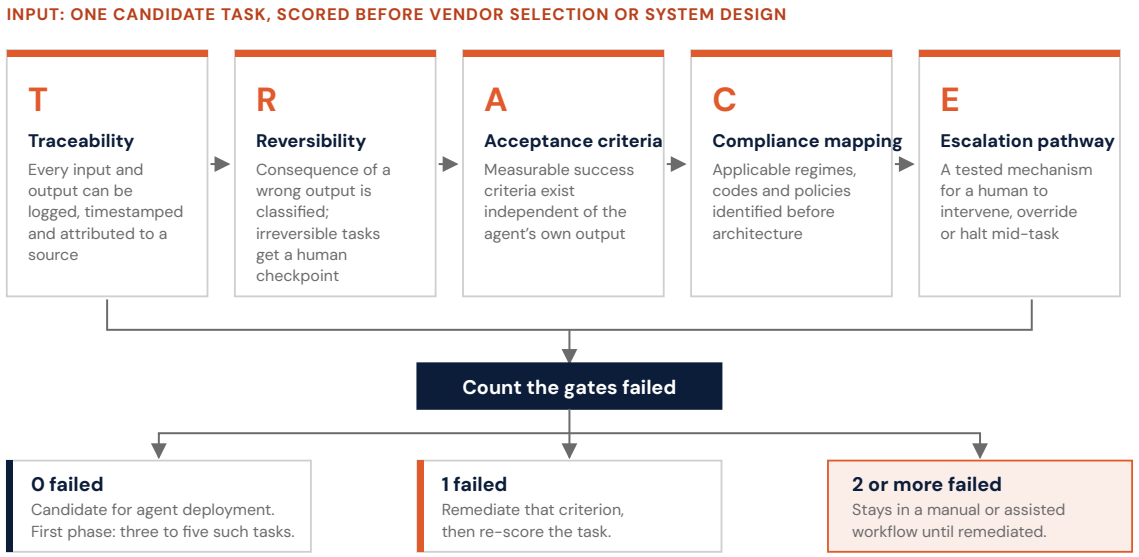
Applied in enterprise AI agent programmes and government agentic AI deployments. Full methodology published in: Beginning Your Journey: Identifying Tasks for Quality, Traceable, Auditable AI Agents.

Organisations moving from AI pilots to autonomous agent deployment consistently fail at the same point: they commission agents before establishing whether a given task is structurally appropriate for delegation. TRACE is a pre-deployment evaluation framework that answers that question before architecture decisions are made.

A task satisfying all five TRACE criteria is a reasonable candidate for agent deployment. A task failing two or more criteria should remain in a manual or assisted workflow until remediated.

REFERENCE DIAGRAM

Five gates applied in sequence to each candidate task before any vendor or architecture decision, and the three routes out depending on how many gates fail.



T

Traceability of Inputs and Outputs

Every input consumed and output produced by the task can be logged, timestamped, and attributed to a specific data source. Tasks drawing on unstructured or unverified data sources should be deprioritised until provenance controls are in place.

R

Reversibility and Risk Tier

The task is classified by consequence of an erroneous output: reversible with no material harm, reversible with cost, or irreversible. Irreversible-consequence tasks require human-in-the-loop checkpoints before any agent autonomy is granted.

A

Acceptance Criteria Definition

The task has measurable, pre-agreed success criteria independent of the agent's own output. A task lacking an external ground truth or validation method is not ready for agent deployment regardless of technical feasibility.

C

Compliance and Regulatory Mapping

The regulatory regimes, sector codes, and internal governance policies that apply to the task are identified before architecture decisions, not retrospectively. For government and regulated environments, this mapping is a prerequisite to scoping.

E

Escalation and Override Pathway

A clear, tested mechanism exists for a human operator to intervene, override, or halt the agent mid-task. Absence of this pathway disqualifies a task from autonomous execution regardless of model performance.

Application sequence: Score each candidate task against all five criteria before any vendor selection or system design. Present the scored task inventory to leadership before procurement. For the first deployment phase, select no more than three to five tasks that satisfy all five criteria, so you can monitor closely before scaling.

CITE THIS FRAMEWORK

VERSION	FIRST PUBLISHED	LAST REVISED	LICENCE
1.1	17 June 2026	12 September 2026	CC BY-NC 4.0

DOI
10.5281/zenodo.22722217

APA | Harvard | Chicago | BibTeX

Kok, T. (2026). TRACE Framework for AI agent task evaluation (version 1.1) [Framework reference]. terencekok.com. https://doi.org/10.5281/zenodo.22722217

Copy citation