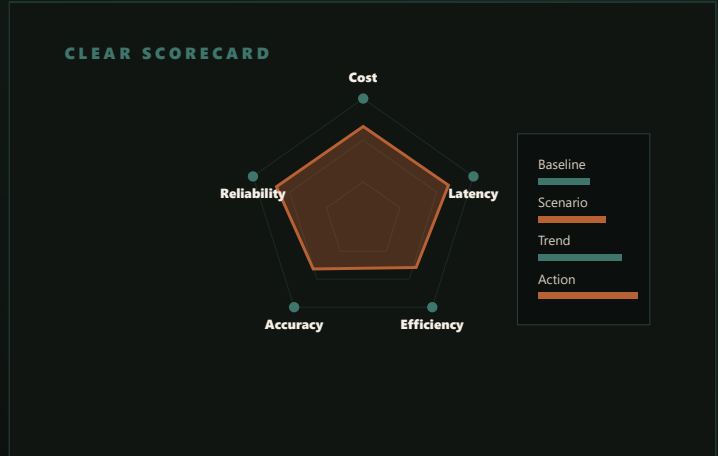


PHANTOM VSA / RESEARCH BRIEF

CLEAR Scorecard Template

A measurement template for proving that an agent is useful, controlled, and worth operating.



01 5 CLEAR dimensions

02 40 scenario seeds

03 1 monthly cadence

A demo is not an operating baseline.

Agent demos often show a successful path without establishing whether the system is economical, fast enough, accurate enough, or reliable under review conditions. CLEAR turns agent quality into an operating cadence. The scorecard should be used from the first diagnostic so HARNEXA can separate impressive demos from agents worth scaling.

01

An agent without a baseline is impossible to improve credibly.

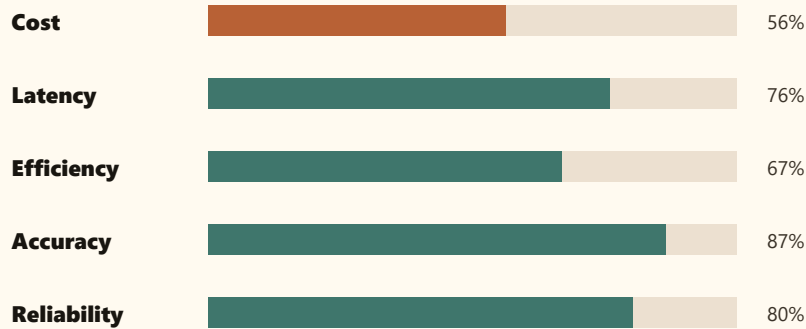
02

Cost and latency are operating facts, not engineering footnotes.

03

Reliability includes correct refusal, hold, and escalation behavior.

CLEAR operating baseline



Monthly review compares the agent against the manual workflow, not against demo theatre.

PWC / AI AGENT SURVEY 2025

Companies adopting agents are looking for measurable productivity value, not only experimentation.

ACCENTURE / FROM EARLY IMPACT TO ENDURING ADVANTAGE

Durable AI impact depends on systemic readiness and end-to-end process redesign, not isolated pilots.

STANFORD HAI / AI INDEX REPORT 2025

AI systems are moving into higher-stakes operating settings, increasing the need for evaluation and governance.

Five score dimensions.

CLEAR converts agent performance into five review dimensions that a buyer can track monthly. Read the artifact as an AgentOps review board: cost, latency, efficiency, accuracy, and reliability become a monthly management rhythm.

Monthly AgentOps scorecard



COST

Track model, tool, infrastructure, and review effort per run so outcome pricing has a real basis.

LATENCY

Measure time to useful output and time to approved terminal state, not only model response time.

ACCURACY

Review groundedness, business-rule correctness, and evidence quality against known scenarios.

RELIABILITY

Record completion, refusal, fallback, approval hold, and recovery behavior across scenario runs.

Separate impressive behavior from scalable performance.

Use the session to lock baselines, scenario coverage, audit samples, and the scores that trigger rollback or reduced autonomy.

INSPECTION QUESTIONS

- What is the current manual baseline for the target workflow?
- Which scenario set represents normal, edge, and forbidden actions?
- What score would justify scaling or reducing autonomy?
- How often will AgentOps review the scorecard after launch?

DO NOT SHIP UNTIL

- Do not report accuracy without evidence and scenario definitions.
- Do not omit cost from quality review.
- Do not use the scorecard only after production launch.

BASELINE	Capture current time, cost, quality, and exception rate for the manual workflow.	Baseline notes and metric definitions.
SCENARIO	Run the agent across representative, edge, and blocked-action scenarios.	Scenario registry with expected outputs and risk class.
CADENCE	Review CLEAR monthly with audit samples and cost bands.	AgentOps report with trend notes and next actions.

Source ledger and measurement boundary

The brief combines external research signals with HARNEXA's implementation boundary: identity, permission, approval, audit, CLEAR evaluation, and no autonomous public execution.

<p>PWC</p> <p>AI agent survey 2025</p> <p>PwC reports that many companies are adopting AI agents, increasing AI budgets, and looking for measurable productivity value.</p> <p>https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-agent-survey.html</p>	<p>ACCENTURE</p> <p>From early impact to enduring advantage</p> <p>Accenture frames systemic readiness and end-to-end process redesign as the binding constraint for durable AI impact.</p> <p>https://www.accenture.com/content/dam/centure/final/accenture-com/document-fy26/q3/Accenture-Impact-to-advantage.pdf</p>	<p>STANFORD HAI</p> <p>AI Index Report 2025</p> <p>Stanford tracks accelerating AI capability, policy attention, governance pressure, and adoption across the global market.</p> <p>https://hai.stanford.edu/ai-index/2025-ai-index-report</p>
<p>RESEARCH STANCE</p> <p>External research is cited as market signal; HARNEXA interpretation is kept separate from source claims.</p>	<p>BUYER ARTIFACT</p> <p>Every report includes a concrete visual artifact that can be inspected in a sales, risk, or architecture review.</p>	<p>OPERATING BOUNDARY</p> <p>CLEAR score, scenario registry, audit sample, monthly review. The asset is for governed review, not autonomous production execution.</p>

HARNEXA BOUNDARY

Technical governance aid only. Not legal advice or compliance certification. HARNEXA AI builds audit-ready foundations for qualified review. No PDF in this library claims EU AI Act certification, legal advice, payment readiness, or autonomous production execution.