

Agent Reliability

Task completion, tool policy, recovery and failure categories.

ILLUSTRATIVE DECISION

Keep tool execution in the test boundary until recovery evidence improves.

AI HEALTH

82

Illustrative

CONFIDENCE

76%

Developing

EVIDENCE

24

Synthetic observations

OPEN ACTIONS

2

Human review

What supports it

- Representative synthetic scenarios
- Quality, reliability and latency signals
- Explicit confidence and evidence window

Known limitations

- No real provider or customer data
- No peak-load or longitudinal evidence
- Cannot prove future safety or performance

Next validation action

- Run the approved representative workload
- Adjudicate open human-review findings
- Record the accountable decision owner

IMPORTANT BOUNDARY

This document demonstrates report structure only. Values, labels and decisions are fictional. A real WAO result applies only to the exact tested workload, provider, model, operation, account, region and evidence window. WAO does not certify or guarantee AI safety, accuracy or production readiness.