

ARBITER

AI Assurance from Formation to Operation

Configure the limits.
Automate the oversight.

Early assessment. Automated monitoring.
Escalation when needed.

Understand the instance before relying on it.

Arbiter is being developed to assess AI instances as they form, monitor for potential drift, hallucination or error, and carry out permitted responses automatically. Teams configure operating policies and escalation rules.

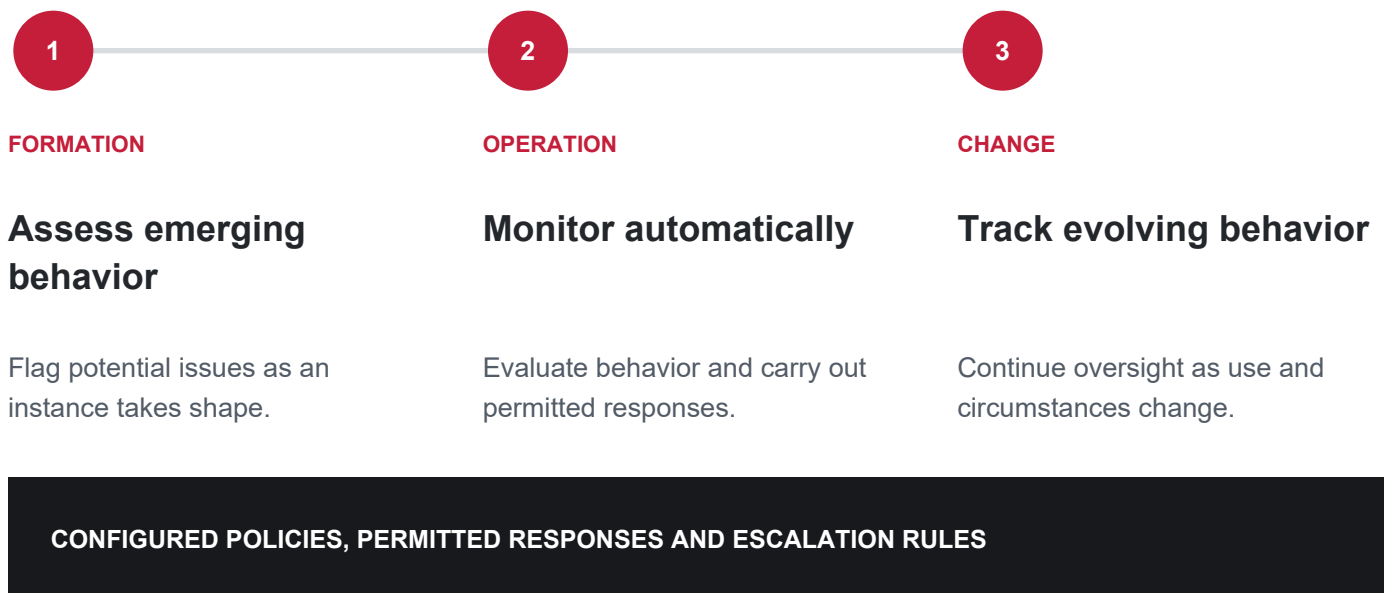
IN DEVELOPMENT / FIRST PLANNED RELEASE: DRIFT SIDECAR

Minimizing added burden and influence is a core design objective. This paper presents the intended customer value and evaluation approach; operational readiness and performance remain to be established.

01 / ASSURANCE ACROSS THE LIFECYCLE

Assess early. Keep monitoring.

Early instance assessment is intended to identify potential problems while an instance takes shape. Once configured, Arbiter is designed to monitor behavior automatically and respond within defined limits.



THE DESIGN PRIORITY

Useful assurance with limited added burden.

The objective is to minimize added processing burden and influence on the instance. An evaluation should measure these effects alongside the usefulness and limits of the findings.

Conceptual customer lifecycle for the intended product.

02 / AN ILLUSTRATIVE CUSTOMER SCENARIO

A policy assistant. Oversight that keeps running.

An internal policy assistant will help employees navigate approval requirements. Before rollout, the team configures acceptable behavior, permitted responses and conditions that require escalation.

THE OPERATING QUESTION

**“Can we keep oversight active
without reviewing every event?”**

WHAT AN ARBITER EVALUATION IS INTENDED TO SUPPORT

01 Configure the use case

Set operating expectations, permitted responses and escalation rules for the assistant.

02 Run routine oversight

Assess the emerging instance and monitor behavior during use, carrying out responses allowed by configuration.

03 Focus attention on exceptions

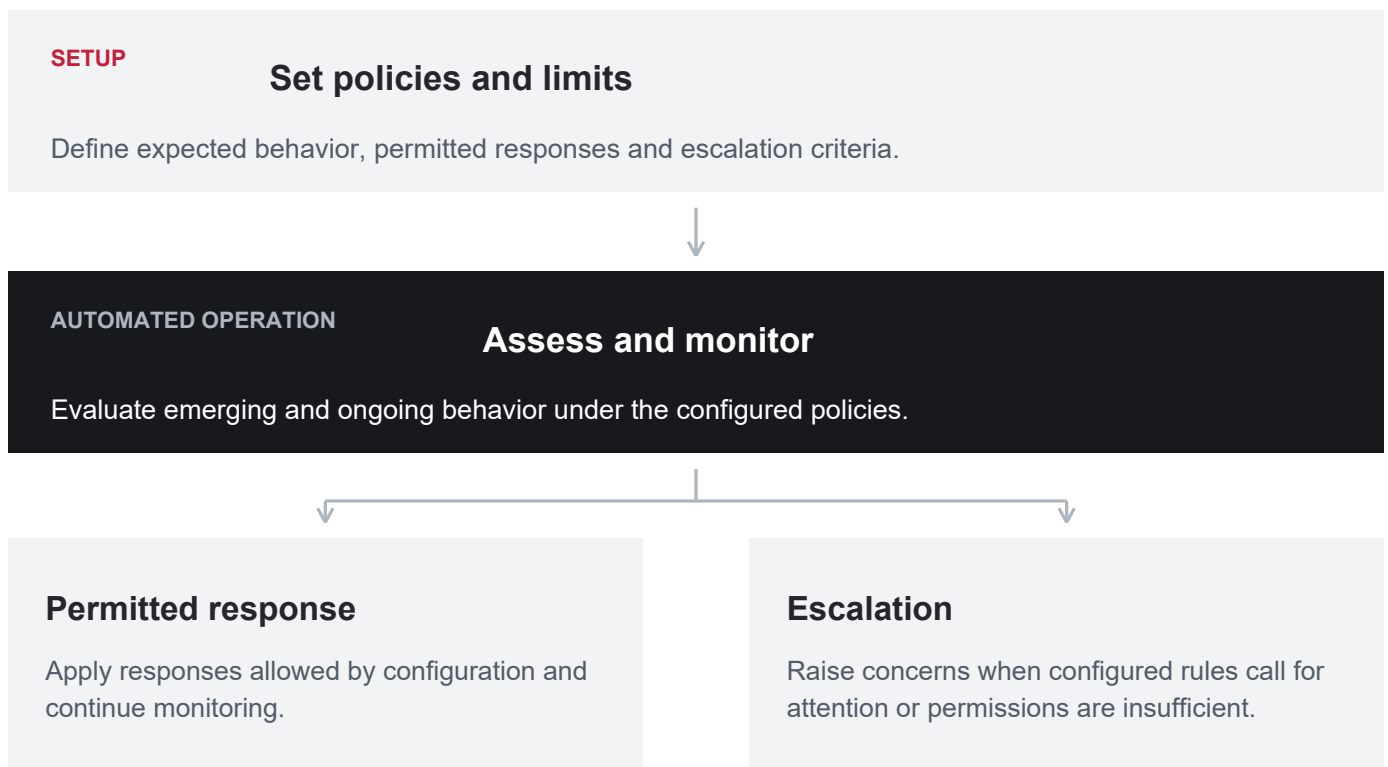
Escalate concerns that need attention, with evidence and uncertainty available for review.

Illustrative intended use. Capabilities remain subject to evaluation.

03 / CONFIGURED AUTOMATION

Configure the limits. Automate routine oversight.

Arbiter is intended to keep routine oversight running under configured policies. Permitted responses can proceed automatically; exceptions trigger escalation with evidence for review.



Changes to approved expectations or expanded permissions still require authorization.

THE BROADER ARBITER PLATFORM

Output verification and multi-model corroboration are part of the broader platform direction. Supporting evidence, disagreement and uncertainty should remain visible. Model agreement alone does not establish correctness.

04 / A FOCUSED FIRST EVALUATION

One assistant. Clear policies. Measured results.

Begin with a bounded use case and clear operating policies. Measure useful findings, permitted automation, escalation quality and the practical cost of oversight.

EVALUATION AREA	QUESTION TO ANSWER
Early usefulness	What potential issues emerged during formation?
Coverage and uncertainty	What was missed or remained inconclusive?
Burden and influence	What added burden or influence did assessment introduce?
Automation and escalation	Which responses ran automatically, and which concerns required attention?
Policy adherence	Did responses stay within configured permissions?

Evaluate the usefulness of findings and responses alongside missed issues, uncertainty, false alarms and the attention required.

Discuss an Arbiter evaluation

spartanxcorp.com/contact