Avahi

MODEL EVALUATION ASSESSMENT

Choose your next model on evidence, not a leaderboard.

We test frontier and open-weight models against your own production tasks and data, through Amazon Bedrock in your AWS account.

WHY NOW

A leaderboard does not know your workload.

Teams running generative AI on one frontier model usually face the next choice after a forcing event:

- Inference cost rising at scale
- A new model claiming similar quality for less
- A renewal or procurement decision
- Pressure to reduce single-vendor lock-in

SCORED FOR EVERY CANDIDATE

Same tasks, same ground truth, every model.

Example candidates

Candidate	Accuracy	Latency	Throughput	Tokens	Cost per task	Cost per successful task
Claude models	✓	✓	✓	✓	✓	✓
OpenAI models	✓	✓	✓	✓	✓	✓
DeepSeek	✓	✓	✓	✓	✓	✓
Kimi	✓	✓	✓	✓	✓	✓
Qwen	✓	✓	✓	✓	✓	✓

Cost per successful task counts only answers that meet your success criteria, so a model that is cheap per call but unreliable does not look like a win. DeepSeek, Kimi, and Qwen run as open-weight models on Amazon Bedrock.

HOW THE ASSESSMENT RUNS

Three stages, one clear recommendation.

Scope and timing are set with each customer.

01
Discovery and test design
We define the workload, pull real tasks and data from production, and set ground truth and success criteria.

02
Evaluation execution
The Avahi model evaluation platform runs every candidate against the agreed test set on Amazon Bedrock.

03
Recommendation and roadmap
A scored comparison, a recommendation, and a plan for moving the chosen model into production.

Optional
Routing and multi-model guidance
Guidance on routing by task, fallback, and multi-model patterns on AWS.

WHAT YOU LEAVE WITH

A full report you can act on.

- Our testing methodology
- Model-by-model results
- A technical recommendation
- Cost projections at your expected volume

SUPPORT

During the assessment, you have direct access to the Avahi team running the evaluation for questions, scope adjustments, and interim findings. Support covers the engagement itself; there is no support period after the final report and recommendation are delivered.

Talk with an expert →

Avahi

© 2026 Avahi

sales@avahi.ai avahi.ai

aws AWS Premier Tier Services Partner

ANTHROP\C Anthropic Select Tier Partner

OpenAI OpenAI Select Tier Partner