Avahi

FIXED-FEE ASSESSMENT

OPENAI ON AMAZON BEDROCK MIGRATION ASSESSMENT

Two plans before you switch.

Moving Claude workloads to OpenAI models on Amazon Bedrock is a cost decision and an engineering decision at once. This fixed-fee assessment answers both before anything moves.

FOR THE CFO AND FINOPS LEAD

A business plan finance can trust

- Total cost of ownership of current Claude spend versus OpenAI models on Amazon Bedrock.
- When existing commitments and reserved or provisioned capacity allow the switch.
- Return on investment where you supply the business metrics.

FOR THE CTO AND AI PLATFORM TEAM

A plan engineering can run

- Target architecture, code and prompt changes, migration waves, risks, and mitigations.
- An evaluation approach that tests output quality on your own tasks.
- Region fit checked up front, since OpenAI model availability varies by AWS Region.

WHAT MOVES WHERE

Every piece gets a destination

Mapped in your target Regions on Amazon Bedrock. Anything without a direct equivalent is flagged with options.

Models and APIs	→	OpenAI models on Amazon Bedrock
Tool use and agent code	→	Amazon Bedrock AgentCore
Retrieval	→	Amazon Bedrock Knowledge Bases
Content filtering	→	Amazon Bedrock Guardrails
Usage and spend	→	AWS Cost Explorer baseline and tracking

HOW IT RUNS

Scope and timing are set with each customer.

01
Discovery
Claude workloads, usage, and billing.

02
Workload mapping
Each component given a destination.

03
Business plan
Cost, commitment timing, and return.

04
Migration plan and readout
Both plans, presented to finance and engineering together.

ENGAGEMENT AND SUPPORT

A fixed fee, with scope agreed before the engagement starts

Support is provided during the assessment only: a named Avahi engagement lead, a kickoff, working sessions with the customer's engineering and finance teams, and answers to questions during business hours. Support ends with the final readout. Carrying out the migration plan is scoped as a separate engagement.

GOOD FIT

Before we start

Claude workloads in production or a late-stage pilot, and access to their usage and billing data.