



The AWS AI Services Map (2026)

AWS documentation and announcements to 2026 | AI services, Amazon Q, Bedrock, SageMaker AI, infrastructure; which service for which need; retired and renamed services

[AWS](#) [AIF-C01](#) [MLA-C01](#) [Services map](#)

Three layers on one infrastructure

The AWS AI stack

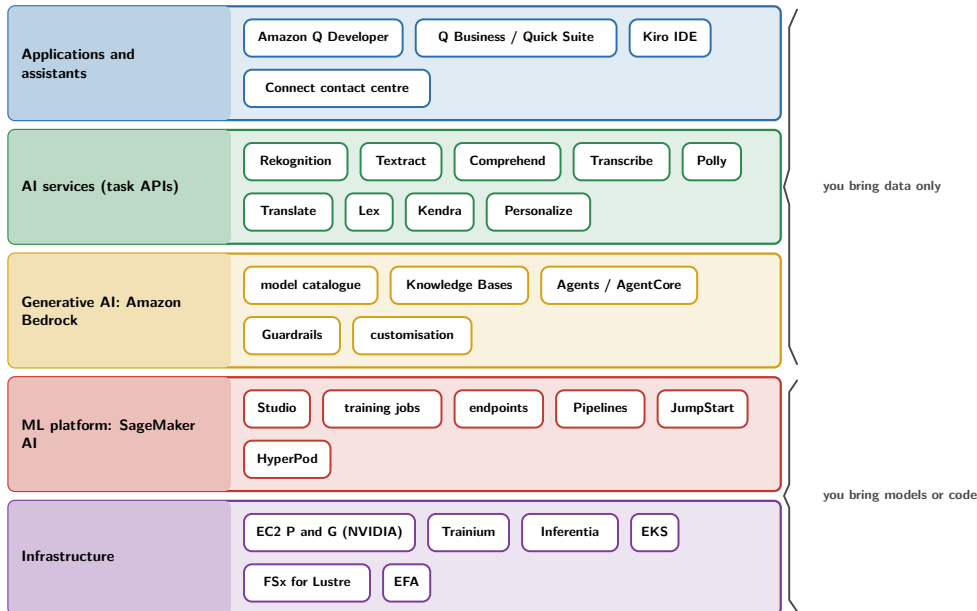


Buy the finished dish when it exists (an AI service), cook with a ready ingredient when it does not (Bedrock), build the kitchen only for a real competitive edge (SageMaker AI).

AI services are single-task APIs billed per page, minute or image; Bedrock is one API to many foundation models billed per token; SageMaker AI runs your containers billed per instance-hour.

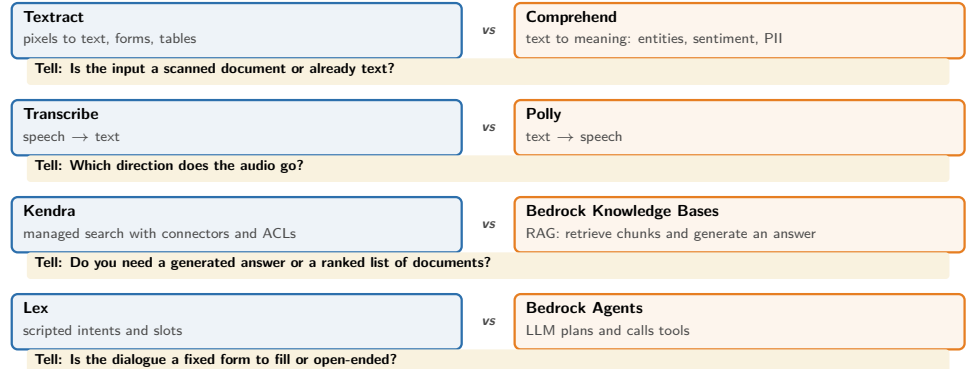


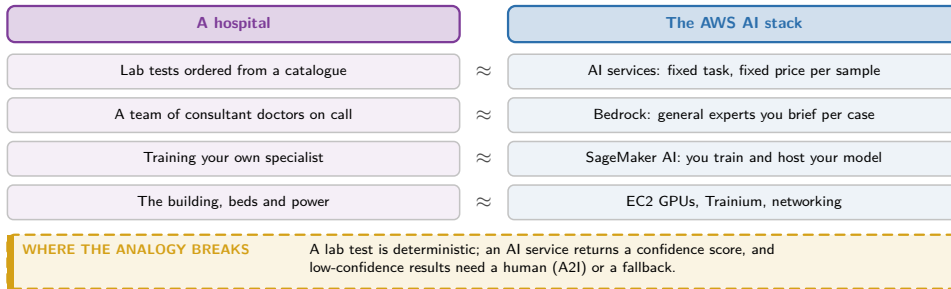
The layers combine: Textract output grounds a Bedrock answer; a SageMaker AI endpoint hosts a fine-tuned open-weight model that Bedrock cannot serve; all run on EC2 GPU, Trainium or Inferentia.



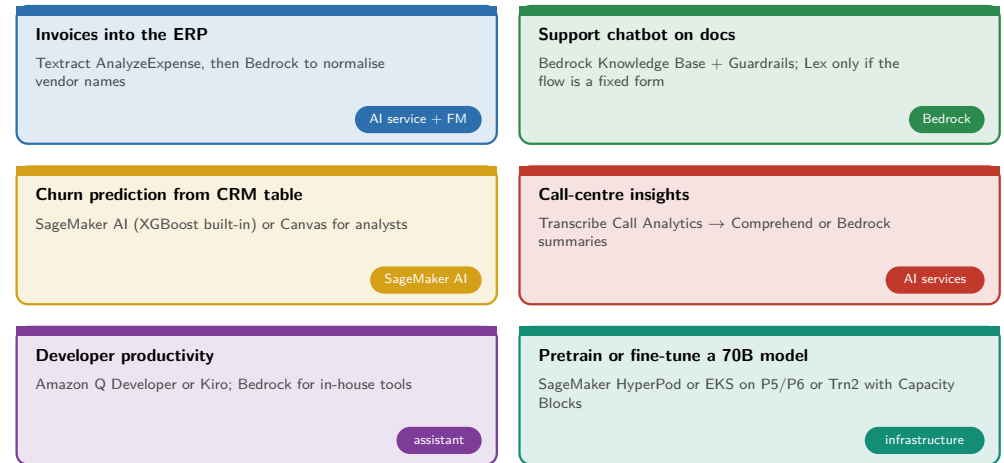
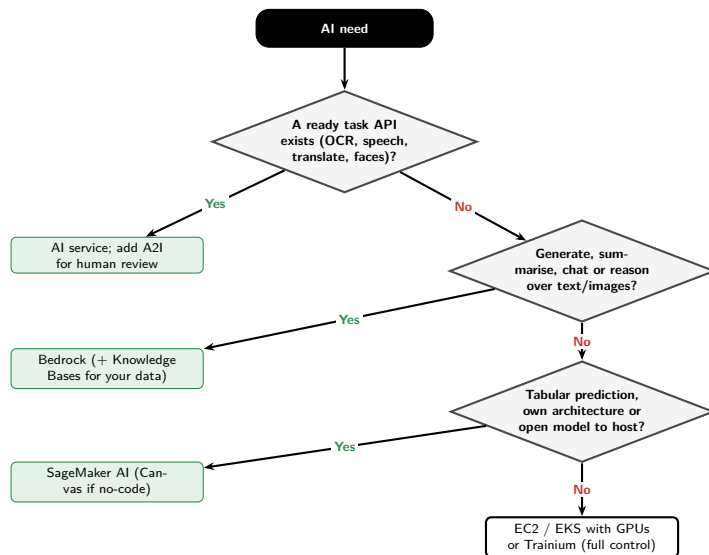
AI services: one task, one API

Service	Need	Key features / APIs	Billing unit
Rekognition	see what is in images and video	DetectLabels, DetectFaces / CompareFaces, DetectModerationLabels, DetectText, Custom Labels, face liveness	image, video minute
Textract	read documents	DetectDocumentText (OCR), AnalyzeDocument (FORMS, TABLES, QUERIES, SIGNATURES), AnalyzeExpense, AnalyzeID, async jobs for multipage PDF	page
Comprehend	understand text	entities, key phrases, sentiment, syntax, language, PII; custom classifier and custom entities; Comprehend Medical for clinical text	unit of 100 characters
Transcribe	speech to text	batch and streaming, speaker diarisation, custom vocabulary and language models, PII redaction, Call Analytics; HealthScribe for clinical notes	audio second
Polly	text to speech	standard, neural, long-form, generative voices; SSML; speech marks for lip-sync	character
Translate	translate text and documents	real-time and batch, custom terminology, Active Custom Translation with parallel data, formality and profanity settings	character
Lex V2	chatbot or voice bot	intents, slots, Lambda fulfilment, Connect integration, generative assisted slot resolution via Bedrock	request
Kendra	enterprise search	connectors (S3, SharePoint, Confluence...), ACL-aware results, GenAI Index used as a Bedrock or Q retriever	index-hour + queries
Personalize	recommendations	datasets (users, items, interactions), recipes such as user personalisation and similar items, real-time campaigns, batch inference	training-hour + requests





© Which service for which need



PROBLEM An insurer receives 20,000 scanned claim forms a month and wants the fields in its database plus a one-paragraph summary per claim. Which services, in which order?

GIVEN

- input: scanned PDFs with checkboxes, tables and handwriting
- low tolerance for wrong amounts; staff can review a few percent

- Extract**
 Textract StartDocumentAnalysis (FORMS, TABLES, QUERIES)
Why: task API, billed per page, returns confidence per field
- Review doubtful fields**
 A2I human loop when confidence < threshold
Why: keeps wrong amounts out of the database
- Summarise**
 Bedrock Converse on the extracted text, guardrail attached
Why: generation needs an FM, not a task API
- Orchestrate**
 S3 event → Step Functions → Lambda per step
Why: serverless, retries, audit trail

ANSWER Textract → A2I (low confidence) → Bedrock summary, orchestrated by Step Functions

CHECK ✓ No SageMaker AI needed: no custom model is trained

EXAM TIP

Exam keywords: **least operational overhead** → a managed AI service or Bedrock; **custom model, full control of training** → SageMaker AI; **no ML expertise, business analyst** → SageMaker Canvas; **human review of low-confidence predictions** → Amazon A2I; **detect PII in text** → Comprehend (or Guardrails sensitive-information filter inside Bedrock).

⚠ Retired, closed and renamed (as of 2026)

- **SageMaker** → **SageMaker AI** renamed Dec 2024; 'Amazon SageMaker' now = unified data + AI platform with SageMaker Unified Studio
- **CodeWhisperer** → **Amazon Q Developer** merged April 2024
- **Amazon Forecast** closed to new customers July 2024; use SageMaker Canvas or SageMaker AI forecasting (Chronos, DeepAR)
- **Lookout for Metrics, Lookout for Vision** end of support in 2025
- **Lookout for Equipment** closed to new customers; check end-of-support date
- **SageMaker Edge Manager** discontinued April 2024; use IoT Greengrass with your own runtime
- **AWS DeepLens** end of life January 2024
- **Lex V1** retired 2025; use Lex V2 APIs
- **SageMaker Studio Classic** replaced by the new Studio (2023) experience
- **Q Business features** moving into Amazon Quick Suite (announced Oct 2025); verify current naming
- **Bedrock, SageMaker AI, Rekognition, Textract, Comprehend, Transcribe, Polly, Translate, Kendra, Personalize** active

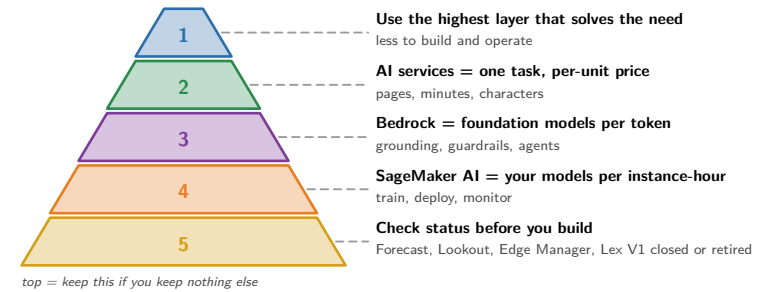
⚠ WARNING

Exams lag product changes: AIF-C01 and MLA-C01 questions may still say 'SageMaker' or mention Forecast. Answer with the concept the question tests; in real projects do not start anything on a service closed to new customers.

Myth or fact?

Statement	True	False
1. Bedrock replaced the AI services. → Task services stay cheaper and more predictable for their task: OCR per page with Textract beats an LLM reading every page; combine both when you need reasoning.	☐	☒
2. SageMaker AI is only for data scientists with notebooks. → It also hosts any container, runs pipelines and serves open-weight LLMs; Canvas serves analysts without code.	☐	☒
3. Amazon Q is a model you can pick and tune. → Q is a set of assistant products running on Bedrock models.	☐	☒
4. Exams may still use names of renamed or closed services. → Answer with the concept the question tests.	☒	☐

🚩 Recap and self-test



Q1 Extract the table rows from a scanned PDF invoice: which API?

A Textract AnalyzeDocument with TABLES (or AnalyzeExpense for invoice fields); asynchronous StartDocumentAnalysis for multipage PDFs.

Q2 Find personal data (names, addresses) in support emails?

A Comprehend DetectPiiEntities (or the Bedrock Guardrails sensitive-information filter when text flows through an FM).

Q3 Moderate user-uploaded images?

A Rekognition DetectModerationLabels.

Q4 Add a human review step when Textract confidence is low?

A Amazon Augmented AI (A2I) human loop.

Q5 Forecast demand for a new AWS customer in 2026?

A Not Amazon Forecast (closed to new customers 2024): SageMaker Canvas forecasting or a model on SageMaker AI.

Q6 What is the billing unit of Polly and Translate?

A Characters processed.

Q7 Kendra vs Bedrock Knowledge Bases in one line?

A Kendra returns ranked documents with ACL filtering; a Knowledge Base retrieves chunks and generates a grounded answer (it can use a Kendra GenAI Index as retriever).

Q8 What happened to CodeWhisperer?

A Folded into Amazon Q Developer in April 2024.