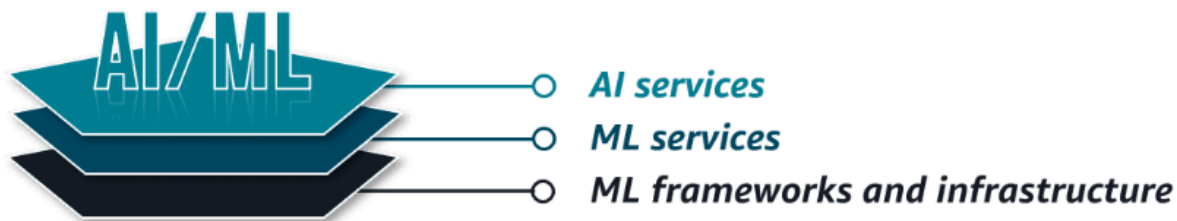**Machine Learning (ML):**
- Subset of AI **(without explicit instructions)**
- Make predictions or decisions based on the patterns it's learned.

AWS AI/ML solutions

The AWS AI/ML stack is composed of the following three tiers of solutions:

- *AI services* - pre-built models that are already trained to perform specific functions

- *ML services* - a more customized approach with Amazon SageMaker AI where you build, train, and deploy your own ML models with fully managed infrastructure

- *ML frameworks and infrastructure* - a completely custom approach to building models using purpose-built chips that integrate with popular ML frameworks

Tier 1: Pre-built AWS AI services: *(Managed)*****
- Pre-built models that are already trained to perform specific functions.

 # Language services;

Amazon Comprehend:
- Uses NLP to **extract key insights** from documents.
- Recognizing key phrases, language, sentiment.
- Use cases: Content classification, customer sentiment analysis, and compliance monitoring

Amazon Polly:
- Converts **text into speech**.
- Use cases: Virtual assistants, e-learning applications, and accessibility enhancements

Amazon Transcribe:

- Converts **speech into text.**
- Speaker identification, custom vocabulary, and real-time transcription.
- Use Cases: Automated subtitling, and metadata generation for media content.

Amazon Translate: - Text translation service.

Computer vision & search services:

Amazon Kendra:
- Uses **NLP to search for answers within large enterprise content.**
- Understands context of a query; can return more precise and relevant answers than just a list of documents with matching keywords.
- Use cases: Intelligent search, chatbots, and application search integration

Amazon Rekognition:
- **Video analysis service.**
- Identify objects, people, text, scenes, and activities within images and videos stored in S3
- Content moderation, identity verification, media analysis, and home automation experiences

Amazon Textract:
- Detects & extracts typed/handwritten text found in documents
- Use cases: Financial, healthcare, and government form text extraction for quick processing

Conversational AI personalization services:

Amazon Lex:
- Add **voice & text conversational interfaces to applications.**
- Uses both NLU and automatic speech recognition (ASR) to create lifelike conversations.
- Use cases: Virtual assistants, natural language search for FAQs, and automated application bots

Amazon Personalize
- Use historical data to build intelligent applications with **personalized recommendations** for customers.
- Use cases: Personalized streaming, product, and trending recommendations

#ML frameworks:
- Software library/tool that provides experienced ML practitioners with **pre-built**, optimized components for building machine learning models.
- PyTorch, Apache M-X Net, and TensorFlow.

Generative AI on AWS:

Amazon SageMaker JumpStart:
- An ML hub with FMs and pre-built ML solutions deployable with a few clicks
- Rapid ML model deployments, Custom fine-tuned solutions, ML experiments and prototypes

Amazon Bedrock: *****Managed*****
- Adapt & deploy FMs from Amazon and leading AI companies (Claude)
- Enterprise-grade generative AI, **Multimodal content generation**, **Advanced conversational AI**

Amazon Q:
- **Interactive AI assistant** that can be integrated with a company's information repositories

Amazon Q Business:
- Answer pressing questions, help solve problems, and take actions using the data and expertise found in company's information repositories.

Amazon Q Developer:
- Code recommendations to accelerate development for coding languages including C#, Java, JavaScript, Python, and TypeScript applications.
- Integrates with multiple IDEs.

AI/ML & Generative AI Services (Exam Cues)

Service	Key Purpose
Amazon SageMaker JumpStart	Pre-built ML models & solutions, fine-tuning, rapid deployment
Amazon Bedrock	Fully managed, multi-model **generative AI** (text, images, etc.) **via single API**
Amazon Q Business	AI assistant for org data — answers, workflows, insights
Amazon Q Developer	AI code generator & reviewer inside IDEs

AWS Data Pipeline Services (I-S-C-P-A):

1. Data Ingestion

Service	Use
Amazon Kinesis Data Streams	Real-time ingestion, massive throughput
Amazon Data Firehose	Near real-time ingestion to S3, Redshift, etc.

2. Data Storage

Service	Use
Amazon S3	Data lake for raw, structured, unstructured
Amazon Redshift	Data warehouse for **structured**/**semi**structured

3. Data Cataloging

Service	Use
AWS Glue Data Catalog	Central **metadata** repository

4. Data Processing

Service	Use
AWS Glue	Fully managed ETL
Amazon EMR	Big data frameworks (Spark, Hadoop, Hive)

5. Data Analysis & Visualization

Service	Use
Amazon Athena	Serverless SQL on S3 and other sources
Amazon Redshift	Frequent high-performance queries
Amazon QuickSight	Interactive dashboards
Amazon OpenSearch Service	Search & real-time monitoring (Visualize data)