Term/Concept	Description
Top-P	0 to 1 value — cumulative probability, balances randomness and accuracy, 1 is least random.
Temperature	Controls randomness—higher values lead to diverse, creative outputs; lower values make results more deterministic.
Epochs	Number of iterations on training a model—more is generally better, but too many can lead to overfitting.
AWS Rekognition	Computer vision/image recognition, used for object detection, facial analysis, and content moderation.
AWS Textract	**OCR** service — converts scanned images into text, structured data extraction from documents.
Amazon Comprehend	Extracts key phrases, entities, sentiment, and PII data from text.
AWS Intelligent Document Processing (IDP)	Automates the processing of unstructured documents (e.g., PDFs, invoices) using Textract, Comprehend, and A2I.
Amazon Lex	Service for **building chatbots** with natural conversational text or voice interfaces.
Amazon Transcribe	**Speech-to-text** service, including audio captioning.
Amazon Polly	**Text-to-speech** service for converting written text into spoken words.
Amazon Kendra	Intelligent document search engine with semantic understanding.
Amazon Personalize	Service for **personalized product recommendations**.
Amazon Translate	**Language translation** service.
Amazon Forecast	Provides time-series **forecasting**, e.g., inventory levels prediction.
Amazon Fraud Detection	Detects fraudulent activity, including online transactions and account takeovers.
Amazon Q Business	Generative AI-powered assistant for enterprise data processing and tasks.
Amazon Macie	**PII data** detection and anonymization service for data security.
SageMaker Clarify	**Bias detection** and explainability for machine learning models.
SageMaker Ground Truth	Provides **human labeling** for **model training** datasets.
Amazon Augmented AI (A2I)	**Human review** for low-confidence predictions during **inference**.
AWS Data Exchange	Access **third-party datasets** securely.
AWS Glue Transformations	ETL transformations like removing duplicates and filling missing values.
Amazon SageMaker JumpStart	Hub with pre-trained models and one-click deployments.
Amazon SageMaker Canvas	No-code tool for building and training machine learning models.
Fine-Tuning	**Adjusting model weights** using **labeled data** to improve task performance.
Domain adaptation fine-tuning	Tailor a model for a **specific domain** like legal or medical using small datasets.
Instruction-based fine-tuning	Fine-tuning a model to **perform better on specific tasks**, e.g., classification.
Continued-Pretraining	Using **unlabelled data** to **expand a model's knowledge base**.
Automatic Model Tuning (AMT)	Automatically tunes hyperparameters to improve model performance.
Data-Drift	Input data changes, but the relationship between inputs and outputs stays the same (e.g., new demographic).
Concept-Drift	The relationship between inputs and outputs changes, meaning the model's learned patterns no longer apply (e.g., new fraud patterns).

Term/Concept	Description
AWS Trusted Advisor	Provides recommendations for cost, performance, and security improvements.
Amazon Inspector	Automated security assessments for application workloads.
AWS PrivateLink	Secure private connections between VPCs and AWS services.
AWS Config	Monitors and records AWS configuration changes for compliance.
AWS CloudTrail	Logs and tracks AWS API calls for auditing.
BedRock GuardRails	Prevents inappropriate foundation model outputs and restricts risky content.
Postgres (Aurora or RDS)	SQL database with vector database support for similarity search.
Amazon DocumentDB	JSON store, MongoDB-compatible with vector database support.
Amazon Neptune	Graph database with vector search capabilities.
Amazon Neptune ML	Uses Graph Neural Networks (GNNs) to predict outcomes from graph data.
Amazon MemoryDB	In-memory database with fast vector search capabilities.
Amazon OpenSearch Service	Search service with vector database support for similarity search.
MSE (Mean Squared Error)	Average squared difference between predicted and actual values, lower MSE indicates better model performance.
RMSE (Root Mean Squared Error)	Square root of MSE, more interpretable; lower RMSE is better.

Confusion Matrix Evaluation Metrics:

Name	When to Use	Example
Precision	When minimizing **false positives** is crucial.	**Spam detection (avoiding false positives)**: Avoiding flagging legitimate emails as spam.
Recall (TPR)	When minimizing **false negatives** is critical, focusing on **identifying all positives**.	**Disease screenings (minimizing false negatives)**: Ensuring that no diseased patients are missed.
False Positive Rate (FPR)	When you need to avoid **false alarms** or incorrect positive predictions.	**Security alarms (avoiding false positives)**: Avoiding alarms for non-threats or normal behavior.
Specificity (TNR)	When minimizing **false positives** and correctly identifying **negatives** is important.	**Non-diseased patients (correct negatives)**: Correctly identifying healthy individuals in medical testing.
Accuracy	For general correctness across both **positives and negatives**.	**Image classification (overall correctness)**: Correctly identifying both cats and dogs in a dataset.
F1 Score	When both **precision and recall** are important, and you want a balance between the two.	**Document classification**: Ensuring correct tagging of documents with minimal errors across both positives and negatives.