

# AI Practitioner – Final Notes

- **RHLF vs A2I vs Ground truth**

- **RHLF vs A2I**

RLHF is a technique used to train AI models using human feedback to refine their behavior, whereas A2I is a service that provides a human review of machine learning predictions to improve model accuracy and reliability

- **Ground Truth vs A2I**

- Amazon SageMaker Ground Truth is a fully managed **data labeling service** that makes it easy to build highly accurate training datasets for machine learning.
    - Amazon Augmented AI makes it easy to build the **workflows using human review** of ML predictions.

- **Feature Store vs Data Wrangler**

Data Wrangler – **Transform** data

Feature store – **Store Features** (or Transformed data labels)

- **Bedrock vs SageMaker Jumpstart**

Amazon Bedrock **provides foundational models** for generative AI applications, whereas Amazon SageMaker JumpStart offers pre-built **solutions** and one-click deployment for various machine learning models (**basically to build models**)

- **AWS Config vs Inspector vs Audit Manager**

**AWS Audit Manager**

- automates continuous audit of AWS usage. It simplifies the process of **assessing risk and compliance** with regulations and industry standards, making it an essential tool for governance in AI systems.
  - **Not continuous monitoring**

**AWS Config** enables you to assess & audit the configurations of your AWS resources. It

- **continuously monitors** and records **config** (AWS resource configurations)
  - allows **automated compliance checking** against **desired** configurations.

This is crucial for governance in AI systems, ensuring that resources remain in compliance with organizational policies and regulatory requirements.

**Amazon Inspector** - Amazon Inspector is an **automated security assessment service** that

- **identifies vulnerabilities and deviations from best practices**, but it is **not primarily focused on continuous monitoring** and compliance of resource configurations.
  - **Used for serverless, EMR, EC2 instances**.

- **Supervised vs Unsupervised vs Semi-Supervised learning algorithms**

- **Supervised** : *Regression, Decision tree, Neural network*

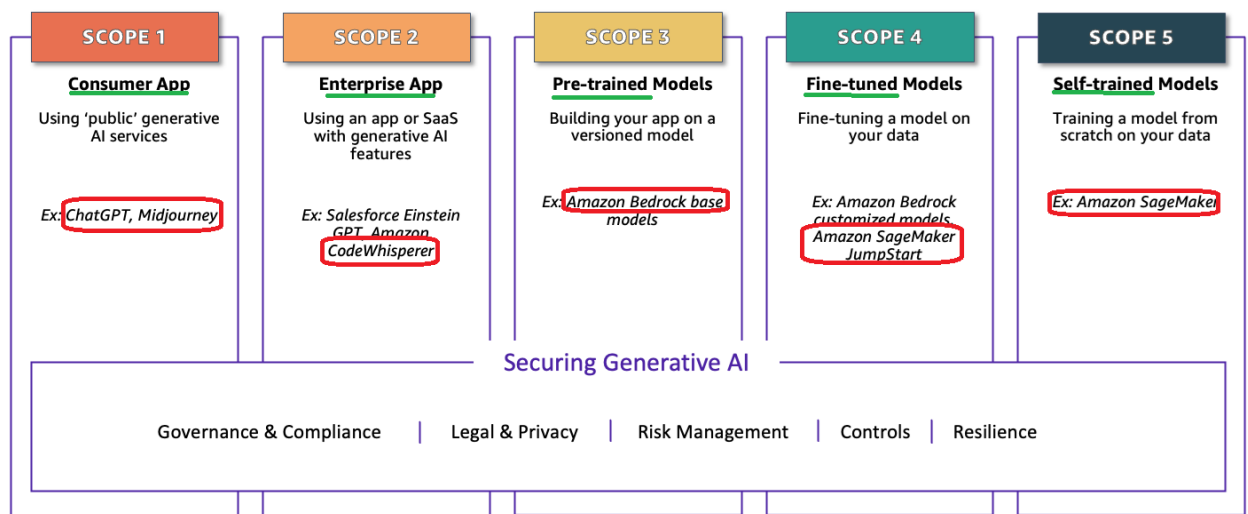
- **Unsupervised:** Clustering, Association rule learning, Probability density, Dimensionality reduction
- **Semi-Supervised:** Fraud identification, Sentiment analysis, Document classification
- **CNNs vs RNNs**
  - CNNs -> single image analysis, RNNs -> video analysis
  - CNN are specifically designed for processing and classifying image data.
  - CNNs are a type of deep learning model particularly well-suited for processing grid-like data, such as images. They are designed to automatically and adaptively learn spatial hierarchies of features from input images.
  - RNNs are designed to handle sequential data, where the order of the data points matters. They are particularly well-suited for time-series data and tasks where temporal dependencies are crucial.
- **Amazon SageMaker Asynchronous Inference deployment types**
  - **Amazon SageMaker Asynchronous Inference:** This option is ideal for
    - requests with large payload sizes (up to 1GB),
    - long processing times (up to one hour)
    - near real-time latency requirements.
  - **SageMaker batch transform:** To get predictions for an entire dataset
  - **SageMaker real-time hosting:** ...
  - **SageMaker Serverless Inference:** For workloads that can tolerate cold starts
- **Amazon Transcribe vs Comprehend**

**Transcribe:** Automatically converts speech to text, including timestamps for each word. This can be used to analyze media content, create meeting notes and subtitles, and more. For example, businesses can use Transcribe to analyze customer support or sales calls to improve their service quality.

**Amazon Comprehend:** Uses NLP to find insights and relationships in text. For example, Comprehend can be used to perform sentiment analysis on text data from Transcribe, returning a percentage for positive, negative, mixed, or neutral sentiments. Comprehend can also extract entities and key phrases
- **Amazon Rekognition vs Textract**
  - **Rekognition can** detect objects, text, and unsafe content, analyze images/videos in any image or video file.
  - **Textract** is a document analysis service that detects and extracts text e.g. printed text, handwriting, structured data (such as fields of interest and their values), and tables from images and scans of documents.
- **Generative AI Security Scoping Matrix**

Let's explore each security discipline and consider how scoping affects security requirements.

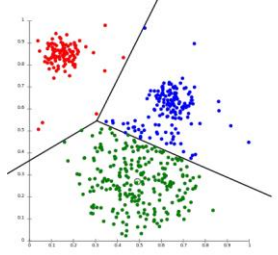
- **Governance and compliance** – The policies, procedures, and reporting needed to empower the business while minimizing risk.
- **Legal and privacy** – The specific regulatory, legal, and privacy requirements for using or creating generative AI solutions.
- **Risk management** – Identification of potential threats to generative AI solutions and recommended mitigations.
- **Controls** – The implementation of security controls that are used to mitigate risk.
- **Resilience** – How to architect generative AI solutions to maintain availability and meet business SLAs.



## • Metrics for model evaluation

Metrics	Purpose	Types of models
Precision, Recall, and F1-Score	Metrics to evaluate model	Classification models
Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared	Metrics to evaluate models	Regression models
Throughput, Latency and Uptime	measure model performance and reliability	All models

- **Algorithms** for model

Supervised (Y/n)	Types of	Algorithms
Unsupervised		<b>k-means</b> 
Supervised	<ul style="list-style-type: none"> <li>• <b>Classification</b></li> <li>• <b>Regression</b></li> </ul>	<b>XGBoost</b> <b>KNN</b>

- **Amazon Bedrock pricing**
  - **On-Demand**
    - **Text model:** every **input token** processed + every **output token** generated
    - **For embeddings models:** Charged for every **input token** processed
    - **For image-generation models:** Charged for **every image generated**.
  - **Batch**
    - (The responses processed and stored in S3 bucket so you can access them later). **50% lower price** compared to on-demand inference pricing
  - **Provisioned Throughput:** Charged by the hour, flexibility to choose between 1-month or 6-month commitment terms. **Mandatory for Only for non-FM (custom)**
  - **Customization:** Model training (total #of tokens processed) x # of hours of use
- **Foundation model parameters:**
  - **Top-k:** Samples tokens with the highest probabilities **until specified number of tokens is reached**.
  - **Top-p** sampling samples tokens with the highest probability scores **until the sum of the scores reaches the specified threshold value (Cumulative probability)**. Top-p sampling is also called nucleus sampling.

## Metrics for fine-tuning LLMs in Autopilot

- **Perplexity loss** measures how well the model can predict the next word in a sequence of text, with lower values indicating a better understanding of the language and context.
- Recall-Oriented Understudy for Gisting Evaluation (**ROUGE**)

Set of metrics used in the field of natural language processing (NLP) and machine learning to evaluate the quality of machine-generated text, such as text summarization or text generation

## Valid model customization methods for Amazon Bedrock?

**NOT RAG** - It allows you to customize a model's responses, not the model itself (When your data changes frequently, like inventory or pricing, it's not practical to fine-tune and update the model while it's serving user queries.)

Foundation models are extremely capable and enable some great applications, but what will help drive your business is generative AI that knows what's important to your customers, your products, and your company. And that's only possible when you supercharge models with your data. Data is the key to moving from generic applications to customized generative AI applications that create real value for your customers and your business.

In this section, we discuss different techniques and benefits of customizing your FMs. We cover how model customization involves further training and changing the weights of the model to enhance its performance.

### Fine-tuning

Fine-tuning is the process of taking a pre-trained FM, such as Llama 2, and further training it on a downstream task with a dataset specific to that task. The pre-trained model provides general linguistic knowledge, and fine-tuning allows it to specialize and improve performance on a particular task like text classification, question answering, or text generation. With fine-tuning, you provide labeled datasets—which are annotated with additional context—to train the model on specific tasks. You can then adapt the model parameters for the specific task based on your business context.

You can implement fine-tuning on FMs with [Amazon SageMaker JumpStart](#) and Amazon Bedrock. For more details, refer to [Deploy and fine-tune foundation models in Amazon SageMaker JumpStart with two lines of code](#) and [Customize models in Amazon Bedrock with your own data using fine-tuning and continued pre-training](#).

### Continued pre-training

Continued pre-training in Amazon Bedrock enables you to teach a previously trained model on additional data similar to its original data. It enables the model to gain more general linguistic knowledge rather than focus on a single application. With continued pre-training, you can use your unlabeled datasets, or raw data, to improve the accuracy of foundation model for your domain through tweaking model parameters. For example, a healthcare company can continue to pre-train its model using medical journals, articles, and research papers to make it more knowledgeable on industry terminology. For more details, refer to [Amazon Bedrock Developer Experience](#).

## Machine Learning governance tools that Amazon SageMaker

- Amazon SageMaker Role Manager
- Amazon SageMaker Model Cards
- Amazon SageMaker Model Dashboard

## Prompt Attacks: Hijacking vs Jailbreaking

- Hijacking involves manipulating an AI system to serve malicious purposes or to misbehave in unintended ways.
- Jailbreaking refers to bypassing the built-in restrictions and safety measures of AI systems to unlock restricted functionalities or generate prohibited content.
- Example of hijacking: where the AI model initially provides a useful response but then diverts to an unethical suggestion (using a cheat tool).
- Example of jailbreaking, where the AI is manipulated into providing information about disabling antivirus software **despite the initial innocent prompt**.

## Deep Racer

- **reinforcement learning** (RL) based
- fully autonomous 1/18th scale race car
- train, evaluate, and tune RL models in the online simulator, deploy their models onto AWS DeepRacer

## SageMaker Models

- **DeepAR** - algorithm i used for forecasting one-dimensional time series data
- **Random Cut Forest** – For detecting anomaly in data
- **Blazing Text** - best suited for text classification and creating word embeddings?

## SageMaker Clarify techniques

- **LIME** - technique for explainability
- **SHAP** (SHapley Additive exPlanations) - technique to measure **the impact of each feature** by evaluating the model's performance with the feature left out.

# AI Use cases

## Computer Vision

field of artificial intelligence that allows computers to interpret and understand digital images and videos. Deep learning has revolutionized computer vision by providing powerful techniques for tasks such as image classification, object detection, and image segmentation.

- **Autonomous driving:** Use computer vision technology to make self-driving cars safer and **more reliable**.
- **Healthcare or medical imaging:** Using computer vision can improve the **accuracy and speed of medical diagnoses**, which leads to better treatment outcomes and increased life expectancy for patients.
- **Public safety and home security :** image and facial recognition can **swiftly identify unlawful entries** or persons of interest, which fosters safer communities and works as a crime deterrent.

## Natural language processing

- **Insurance:** Insurance companies can use NLP to extract policy numbers, expiration dates, and other personal information.
- **Telecommunication:** Telecom companies use NLP to analyze customer text messages and suggest personalized recommendations.
- **Education:** In the education industry, students use Q&A chatbots to address questions.

## Intelligent document (ID) processing

- **Financial Services or lending:** Financial services use IDP to extract important information from mortgage applications to accelerate customer response time. It also helps with the underwriting process by identifying incomplete loan packages, tax forms, pay stubs, and other missing data.
- **Legal:** IDP, along with other applications such as optical character recognition (OCR) and NLP, helps eliminate the manual effort of processing documents such as contractual documents, agreements, court filings, and legal dockets.
- **Healthcare:** expedite business quickly and accurately by processing various document types, such as claims and doctor's notes.

## Fraud detection

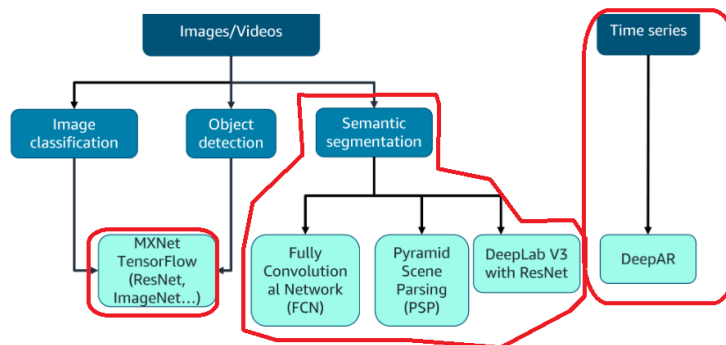
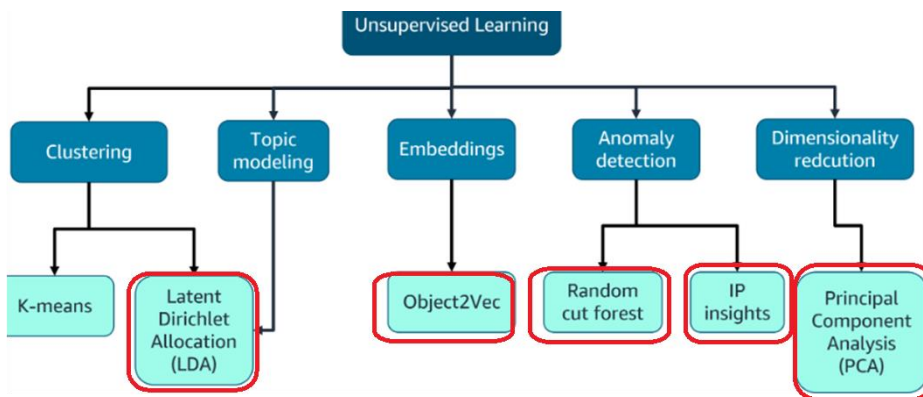
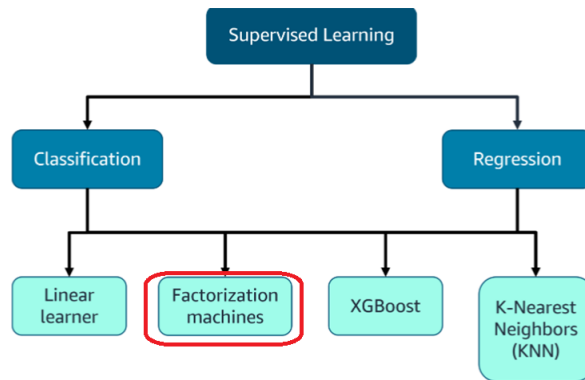
- **Financial Services:** use fraud detection for identity verification, payment fraud detection, transaction surveillance, and anti-money laundering (AML) sanctions.
- **Retail:** Fraud detection systems in the retail industry protect businesses from financial losses, safeguard customer accounts and data, and maintain trust and confidence in online transactions.
- **Telecommunication:**

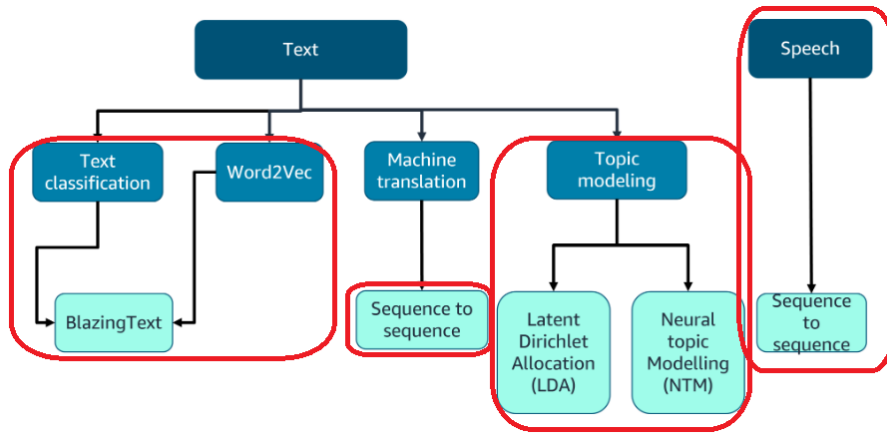
The telecommunication industry uses fraud detection to identify any fraudulent activities in any of the following areas:

- Telecom
  - Roaming, premium rate service, and subscription fraud
- Online
  - New account fraud, claims processing fraud, and promotion abuse
- Retail
  - Credit card and online retail fraud

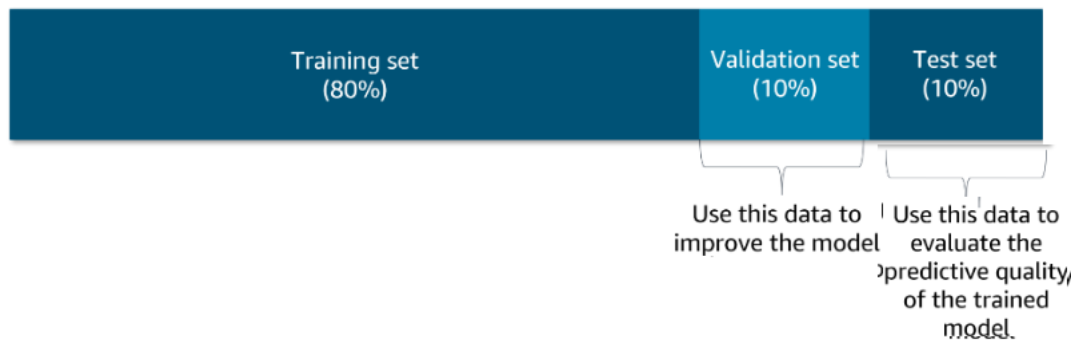


## SageMaker Algorithms





## ML Model training



## ML Confusion Matrix

		Predicted class	
		<i>P</i>	<i>N</i>
Actual class	<i>P</i>	<div>True Positive (TP)</div>	<div>False Negative (FN)</div>
	<i>N</i>	<div>False Positive (FP)</div>	<div>True Negative (TN)</div>

Precision	Recall
$\frac{TP}{TP + FP}$	$\frac{TP}{TP + FN}$

# Optimizing FMs (for Gen AI)

## RAG

### Storing vectors

The core function of vector databases is to compactly store billions of high-dimensional vectors representing words and entities. Vector databases provide ultra-fast similarity searches across these billions of vectors in real time.

The most common algorithms used to perform the similarity search are k-nearest neighbors (k-NN) or cosine similarity.

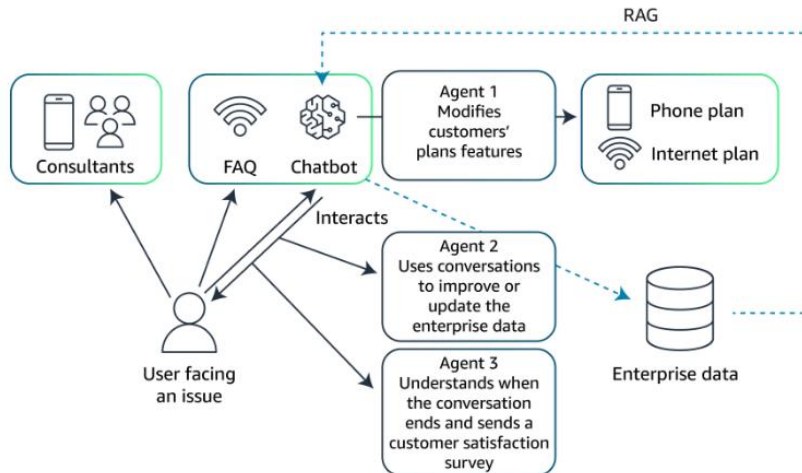
Amazon Web Services (AWS) offers the following viable vector database options:

- Amazon OpenSearch Service (provisioned)
- Amazon OpenSearch Serverless
- Amazon RDS for PostgreSQL (using pgvector extension)
- Aurora PostgreSQL-Compatible Edition using pgvector extension)
- Amazon Kendra

## Agents

Agents can serve different roles in a generative AI application, such as the following:

- **Intermediary operations:** Agents can act as intermediaries, **facilitating communication between the generative AI model and various backend systems.** The generative AI model handles language understanding and response generation.
- **Actions launch:** Agents can be used to run a **wide variety of tasks.** These tasks might include adjusting service settings, processing transactions, retrieving documents, and more. These actions are based on the users' specific needs understood by the generative AI model.
- **Feedback integration:** Agents can also contribute to the AI system's learning process by **collecting data on the outcomes of their actions.** This feedback helps refine the AI model, enhancing its accuracy and effectiveness in future interactions.



## Evaluate Results

### Human evaluation

Human evaluation involves real users interacting with the AI model to provide feedback based on their experience. This method is particularly valuable for **assessing qualitative aspects** of the model, such as the following:

- **User experience:** How intuitive and satisfying is the interaction with the model ?
- **Contextual appropriateness:** Does the model respond in a way that is contextually relevant and sensitive to the nuances of human communication?
- **Creativity and flexibility:** How well does the model handle unexpected queries or complex scenarios that require a nuanced understanding?

### Benchmark datasets

These datasets consist of **predefined datasets and associated metrics that offer a consistent, objective means** to measure model performances. This might include the following:

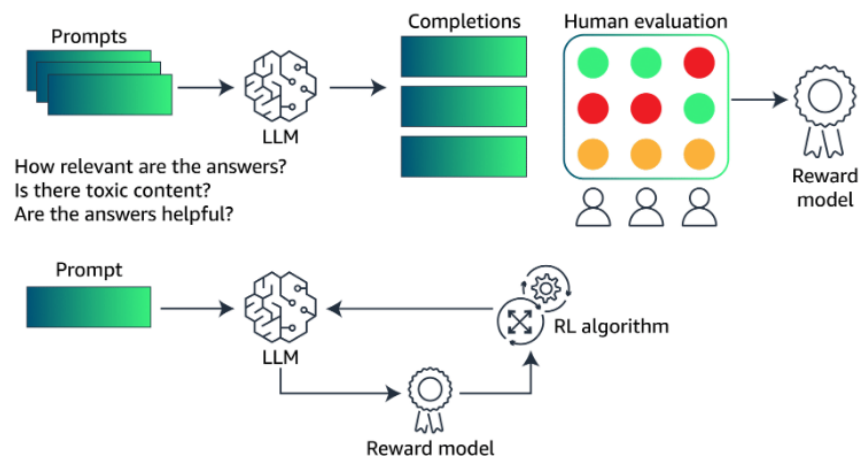
- **Accuracy:** How accurately does the model perform specific tasks according to predefined standards?
- **Speed and efficiency:** How quickly does the mode generate responses and how does this impact operational efficiency?
- **Scalability:** Can the mode maintain its performance as the scale of data or number of users increases?

**When to use Both:** Benchmark datasets are particularly useful for **initial testing phases to ensure that the model meets certain technical specifications before it is put through more subjective human evaluations**. They are also essential for comparing performance across different models or different iterations of the same model.

## Fine Tuning

- **Instruction tuning:** This approach involves retraining the model on a new dataset that consists of prompts followed by the desired outputs. This method is particularly useful for improving the model's ability to understand and execute user commands accurately, making it highly effective for interactive applications like virtual assistants and chatbots.
- **Reinforcement learning from human feedback (RLHF):** This approach is a fine-tuning technique where the model is initially trained using supervised learning to predict human-like responses. Then, it is further refined through a reinforcement learning process, where a reward model built from human feedback guides the model toward generating more preferable outputs.

### RLHF



RLHF refers to the improvement of the model by learning from feedback, such as ratings, preferences, demonstrations, helpfulness, or toxicity, provided by humans. **RLHF is used**

- during the pretraining phase of the model
- but can also be used to fine-tune the model

## Adaptive Approaches

- **Adapting models for specific domains**
- **Transfer learning:** This approach is a method where a model developed for one task is reused as the starting point for a model on a second task. This method is highly efficient in using learned features and knowledge from the general training phase and applying them to a narrower scope with less additional training required.
- **Continuous pretraining:** This approach involves extending the training phase of a pre-trained model by continuously feeding it new and emerging data.

# Responsible AI

## Bias types

- **Data bias:** When training data is biased or underrepresents certain groups, the resulting model may exhibit biases in its predictions or decisions.
- **Algorithm bias:** The algorithms and models used in AI systems can introduce biases, even if the training data is unbiased. This can happen due to the inherent assumptions or simplifications made by the algorithms, in particular for underrepresented groups, or due to machine learning models optimizing for performance, not necessarily for fairness.
- **Interaction bias:** Biases can also arise from the way humans interact with AI systems or the context in which the AI is deployed. For example, if an AI system for facial recognition is primarily tested on a certain demographic group, it may perform poorly on other groups.
- **Bias amplification:** AI systems can amplify and perpetuate existing societal biases, if not properly designed and monitored. This can lead to unfair treatment or discrimination against certain groups, even if it was not intentional.

## Core dimensions of responsible AI

- **Fairness**
- **Explainability:** Humans must understand how models are making decisions and address any issues of bias, trust, or fairness.
- **Privacy and security**
- **Transparency:** Communicate information about an AI system
- **Veracity and robustness:** ensure AI system operates reliably, even with unexpected situations, uncertainty, and errors.
- **Governance:** Processes used to define, implement, and enforce responsible AI practices.
- **Safety:** Refers to the development of algorithms, models, and systems in responsible, safe, and beneficial way for individuals and society as a whole
- **Controllability:** How to monitor and guide an AI system's behavior to align with human values and intent. It involves developing architectures that are controllable, so that any unintended issues can be managed and addressed.

## Core dimensions of responsible AI

- Define application use case narrowly
- Choosing a model based on performance
- Choosing a model based on sustainability concerns (energy consumption, resources utilization, etc.)

## Responsible preparation of data

- **Balancing datasets: To remove bias**
- **Data curation**
  - Preprocess the data to ensure it is accurate, complete, and unbiased. Techniques such as data cleaning, normalization, and feature selection can help to eliminate biases in the dataset.
  - Use data augmentation techniques to generate new instances of underrepresented groups. This can help to balance the dataset and prevent biases towards more represented groups.
  - Regularly audit the dataset to ensure it remains balanced and fair. Check for biases and take corrective actions if necessary.

## Transparency vs Explainability

- **Transparency: HOW** a model makes decisions.
  - This helps to provide accountability and builds trust in the AI system.
  - Transparency also makes auditing a system easier.
- **Explainability: WHY** the model made the decision that it made.
  - It gives insight into the limitations of a model.
  - This helps developers with debugging and troubleshooting the model.
  - It also allows users to make informed decisions on how to use the model.

## AWS tools for explainability

### SageMaker Clarify

SageMaker Clarify is integrated with SageMaker Experiments to provide scores detailing which features contributed the most to your model prediction on a particular input for tabular, NLP, and computer vision models.

### SageMaker Autopilot

Amazon SageMaker Autopilot uses tools provided by SageMaker Clarify to help provide insights into how ML models make predictions. These tools can help ML engineers, product managers, and other internal stakeholders understand model characteristics.



# Security and Compliance in AI

## Defense in depth

- **Data protection**
  - **Data at rest:** Ensure that all data at rest is encrypted with AWS KMS or customer managed key. Make sure all data and models are versioned and backed up using S3 versioning.
  - **Data in transit:** Protect all data in transit between services using AWS Certificate Manager (ACM) and AWS Private Certificate Authority (AWS Private CA). **Keep data within virtual private clouds (VPCs) using AWS PrivateLink.**
- **Identity and access management**
- **Application protection:** AWS Shield + Amazon Cognito+ Others
- **Network and edge protection:** VPC + AWS WAF
- **Infrastructure protection:** IAM + network ACLs
- **Threat detection and incident response**
  - **Threat detection:** AWS **Security Hub** + Amazon **GuardDuty**
  - **Incident response :** AWS Lambda + Amazon EventBridge

## How AI standards compliance differs from traditional software

- **Complexity and opacity:** AI systems, especially large language models (LLMs) and generative AI, can **be highly complex with opaque decision-making processes.**
- **Dynamism and adaptability:** AI systems are **often dynamic and can adapt and change over time,** even after deployment. This makes it difficult to apply static standards, frameworks, and mandates.
- **Emergent capabilities:** Unexpected or **unintended capabilities that arise as a result of complex interactions** within the AI system. **In contrast to capabilities that are explicitly programmed or designed.**
- **Unique risks:** AI poses **novel risks, such as algorithmic bias, privacy violations, misinformation, and AI-powered automation** displacing human workers. Traditional requirements might not adequately address these.
- **Algorithmic bias** refers to the systematic errors or unfair prejudices that can be introduced into the outputs of AI and machine learning (ML) algorithms. The following are some examples:
  - **Biased training data:** If the data used to train the AI model is not representative or contains historical biases, the model can learn and perpetuate those biases in its outputs.
  - **Human bias:** The biases and assumptions of the human developers and researchers who create the AI systems can also get reflected in the final outputs.
- **Algorithm accountability:** Idea that algorithms, especially those used in AI systems, should be transparent, explainable, and subject to oversight and accountability measures.

# Amazon Kendra

Supports data search across many sources, but **no DynamoDB**

- S3,
- SharePoint,
- Salesforce
- ServiceNow
- RDS databases **(no DynamoDB)**
- Microsoft OneDrive, Google Drive
- Etc.

# Amazon Textract

**Q: What document formats does Amazon Textract support?** PNG, JPEG, TIFF, and PDF

**Q: What APIs does Amazon Textract offer?**

- **Detect Document Text API:** OCR and key-value pair detection
- **Analyze Document API:** Detect printed text, handwriting, fields, values, their relationships, tables, and other entities within a document along with their associated confidence scores.
- **Analyze Expense API :** Use normalized key names and column headers when extracting data from invoices and receipts so that downstream applications can easily compare output from many documents.
- **Analyze ID API** understands the context of identity documents such as **U.S. passports and driver's licenses** without the need for templates or configuration.

# Amazon Comprehend

Entity, Classification, Sentiment, Toxicity, PII, Syntax, Phrase, Syntax, Language)

- Custom Entity Recognition:
- Custom Classification API:
- **Entity Recognition API**
- **Sentiment Analysis API**
- Targeted Sentiment:
- PII Identification and Redaction:
- Comprehend toxicity detection
- **Prompt Safety Classification:**
- Keyphrase Extraction API
- Comprehend Events

- **Language Detection API**
- **Comprehend Syntax API**
- **Topic Modeling** (tagging)
- **Multiple language support:** German, English, Spanish, Italian, ...

## Amazon Rekognition

(Label, Text, Face, PPE)

- **Labels**
  - **DetectLabels API.**
- **Faces**
  - **DetectFaces API / CompareFaces API CreateCollection API**
  - **IndexFaces API:**
  - **DeleteFaces API:**
  - **DisassociateFaces API:**
  - **SearchUsersByImage API/ SearchUsers API/ FaceID API /Analyze ID API:**
- **Celebrity**
  - **RecognizeCelebrities API:**
- **Text Detection**
  - **DetectText API :**
  - **StartTextDetection and GetTextDetection APIs:**
- **PPE Detection**
  - **DetectProtectiveEquipment API:**
- **Video APIs**

# Amazon Q in QuickSight

## SPICE & AutoGraph

- **SPICE** : QuickSight is built with SPICE—a Super-fast, Parallel, In-memory Calculation Engine.
- **AutoGraph** : QuickSight has an innovative technology called **AutoGraph** that allows it to **select the most appropriate visualizations** based on the properties of the data, such as cardinality and data type.

## Amazon Bedrock

### Retrieval augmented generation (RAG) - KB chunking

Knowledge Bases for Amazon Bedrock **provides 3 options** to chunk text before converting it to embeddings.

1. **Default option:** Knowledge Bases for Amazon Bedrock automatically splits your document into chunks each containing 200 tokens, ensuring that a sentence is not broken in the middle. If a document contains less than 200 tokens, then it is not split any further. An overlap of 20% of tokens is maintained between two consecutive chunks.
2. **Fixed size chunking:** In this option, you can specify the maximum number of tokens per chunk and the overlap percentage between chunks for Knowledge Bases for Amazon Bedrock, so your document will be automatically split into chunks, **ensuring that a sentence is not broken in the middle**.
3. **Create one embedding per document:** Amazon Bedrock creates one embedding per document. This option is suitable if you have preprocessed your documents by splitting them into separate files and **do not want Amazon Bedrock to further chunk your documents**.

# AI

## Metrics for model evaluation

Metrics	Purpose	Types of models
Precision, Recall, and F1-Score	Metrics to evaluate model	Classification models
Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared	Metrics to evaluate models	Regression models
Throughput, Latency and Uptime	measure model performance and reliability	All models

## SageMaker Clarify techniques

- **LIME** - technique for explainability
- **SHAP** (SHapley Additive exPlanations) - technique to measure the impact of each feature by evaluating the model's performance with the feature left out.

## AI Use cases

### Computer Vision

- **Autonomous driving:**
- **Healthcare or medical imaging**
- **Public safety and home security**

### Natural language processing

- **Insurance:**
- **Telecommunication:** analyze customer text messages and recommendations.
- **Education:** Use Q&A chatbots to address questions.

### Intelligent document (ID) processing

- **Financial Services or lending:** extract important information from mortgage applications, identifying incomplete loan packages, tax forms, pay stubs, and other missing data.
- **Legal:** With OCR and NLP, automated processing of documents such as contractual documents, agreements, court filings, and legal dockets.
- **Healthcare:** expedite business quickly and accurately by processing various document types, such as claims and doctor's notes.

### Fraud detection

- **Financial Services | Retail | Telecommunication:**

## Transparency vs Explainability

- **Transparency:** **HOW** a model makes decisions.
  - This helps to **provide accountability and builds trust** in the AI system.
  - Transparency also **makes auditing a system easier**.
- **Explainability:** **WHY** the model made the decision that it made.
  - It gives **insight into the limitations of a model**.
  - This helps developers with debugging and troubleshooting the model.
  - It also allows users to make informed decisions on **how to use the model**.

## How AI standards compliance differs from traditional software

- **Complexity and opacity:** can **be highly complex with opaque decision-making processes**.
- **Dynamism and adaptability:** AI systems are **often dynamic and can adapt** so it difficult to apply static standards, frameworks, and mandates.
- **Emergent capabilities:** Unexpected or **unintended capabilities can arise ...**
- **Unique risks:** AI poses **novel risks, such as algorithmic bias, privacy violations, misinfo..**
- **Algorithmic bias:**
  - **Biased training data:**
  - **Human bias:**
- **Algorithm accountability**cccccccccccccccccccccccccccccccccccc

# AWS Services

- **AWS Config vs Inspector vs Audit Manager**

## **AWS Audit Manager**

- automates continuous audit of AWS usage. It simplifies the process of **assessing risk and compliance** with regulations and industry standards, making it an essential tool for governance in AI systems.
- **Not continuous monitoring**

**AWS Config** enables you to assess & audit the configurations of your AWS resources. It

- **continuously monitors** and records **config** (AWS resource configurations)
- allows **automated compliance checking** against **desired** configurations.

This is crucial for governance in AI systems, ensuring that resources remain in compliance with organizational policies and regulatory requirements.

**Amazon Inspector** - Amazon Inspector is an **automated security assessment service** that

- **identifies vulnerabilities and deviations from best practices**, but it is **not primarily focused on continuous monitoring** and compliance of resource configurations.
- **Used for serverless, EMR, EC2 instances**.

# Misc AI

## Modify the behavior of the algorithm or improve performance of model

- **Hyperparameter tuning** is a method to adjust the behavior of an ML algorithm. You can make changes to an ML model by using hyperparameter tuning to modify the behavior of the algorithm.
- **Feature engineering** is a method to select and transform variables when you create a predictive model. Feature engineering includes feature creation, feature transformation, feature extraction, and feature selection. Feature engineering enhances the data by increasing the number of variables in the training dataset to ultimately improve model performance.

## Identify data quality issues, model quality issues, bias drift, and feature attribution drift.

Model monitoring is a component of the ML lifecycle that captures data and compares the data to the training data. You can use model monitoring to identify data quality issues, model quality issues, bias drift, and feature attribution drift.



## Note

- Gen AI vs Traditional Models
  - If asking about **generating or developing new content**, use **Gen AI models** or Gen AI task (FM or custom)
  - **predicting or classifying a target variable**, use Traditional models (e.g. identify sentiment, predict a value, etc.)
- **Instruction-based fine-tuning** - uses labeled examples that are formatted as **prompt-response pairs** and that are phrased as instructions
- **Regression vs Classification** – Regression generates a number (or yes/no)
- **Domain adaptation fine-tuning** vs **Instruction-based fine-tuning**:

Domain adaptation fine-tuning can expand an LLM's knowledge of a given domain only, whereas **Instruction-based fine-tuning (using labeled data) can perform any specific task like participate in multi-turn chatbot, be creative, etc.**

- **READ the question clearly, did they ask to evaluate models (EVALUTATION), or show the performance of a model (MEASURE), etc.**
- **Continued pre-training of FMs vs RAG vs Instruction-based fine-tuning**
  - Continued pre-training of FMs can help the model understand industry-specific terminology **using unlabeled datasets**
  - Instruction-based fine-tuning uses **labeled data set** to train an FM
  - RAG only indexes internal data and industry documentation **as a RAG source to search to include in a prompt**, it **does not improve the model's ability to incorporate industry-specific terminology in future responses**
- **Textract** and **Kendra** can do respective search, but **can't generate/summarize text.**
- **AWS Vector Search Databases :**
  - Amazon OpenSearch Service (**default**).
  - RDS for PostgreSQL (**only RDS**).
  - Amazon Neptune ML.
  - Amazon MemoryDB
  - DocumentDB (with MongoDB compatibility)



- **Transfer learning** involves adapting an existing model for a specific application, such as fine-tuning a model to understand a new language.
- **ROUGE -L vs ROUGE-N**
  - **ROUGE-L** uses the longest common subsequence to evaluate the coherence and order of the narrative in the generated text. Indicate similarity between the two sequences.
  - **ROUGE-N** measures the number of matching n-grams between the model-generated text and a human-produced reference.
- **SHAP vs LIME**

LIME and SHAP are both methods for explaining the logic behind ML model predictions

- **SHAP** (SHapley Additive exPlanations) is the technique used by SageMaker Clarify to measure the impact of each feature by evaluating the model's performance with the feature left out.
  - **LIME** creates a local explanation by fitting a simple model around a prediction
- **In-transit Security between S3 and Bedrock**
    - Configure Bedrock to use a VPC to keep data transfers within the AWS network.
    - Set up AWS PrivateLink to create a private connection between Bedrock and S3.