Fundamentals of AI and ML

Traditional Programming vs. AI

- **Traditional Programming:** Rule-Based, Deterministic, Transparent (logic).
- **AI:** Data-driven, adaptable, less transparent.

AI Hierarchy

- AI -> ML -> DL->Gen AI.

Foundation Models

- Large general-purpose pre-trained models that are expensive to create.

Training Data

- Consists of Input Variables and Target Variables (Correct label).

Model Fit Patterns

- **Underfitting:** The model performs poorly on both training data and new data.
- **Overfitting:** The model performs well on the training set but poorly on unseen data.
- **Balanced:** The model performs well on both training and new data.
    - A balanced fit does not necessarily mean the model is fair, as it can reflect biases present in the training data.

Types of Machine Learning

- **Supervised Learning:** Uses labeled data.
    - **Classification:** Predicts a category.
        - **Binary:** Two possible outcomes.
        - **Multiclass:** More than two possible outcomes.
    - **Regression:** Predicts continuous values.
- **Unsupervised Learning:** Uses unlabeled data.
    - **Clustering:** Groups similar data points together.
    - **Anomaly Detection:** Identifies rare items or events.
- **Semi-supervised Learning:** Uses a mix of partially labeled data.
- **Self-supervised Learning:**
    - Uses information within the data to learn patterns and features without explicit labels.
    - The model generates its own labels from the data.

- **Reinforcement Learning:** An agent learns by interacting with an environment to achieve a goal.

Data Types

- **Structured Data:** Tabular data (e.g., spreadsheets, databases).
- **Unstructured Data:** Data without a predefined model (e.g., text, images, audio).

The Machine Learning Pipeline

Data Processing Steps

1. **Generate Data**
2. **Fetch**
3. **Clean**
4. **Prepare**

Exploratory Data Analysis (EDA)

- The process of identifying patterns and anomalies in data before model training.
- **Correlation Matrix:** A tool used in EDA.
    - Positive correlation means that as one variable increases, the other decreases, and vice versa.

Feature Engineering

- The process of using domain knowledge to create features that make ML algorithms work.
- **Remove Unnecessary Features (Feature Engineering)**
- **Feature Extraction:** Create new columns from existing data.
- **Dimensionality Reduction:** Combining features to reduce the number of variables.
- **Categorical Encoding:** Convert categorical (text) values to numbers.
- **Normalization:** Scaling data to a range between 0 and 1.
- **Standardization:** Rescaling data to have a mean of 0 and a standard deviation of 1 by calculating the mean and standard deviation.

Training the Model

- **Train and Tune:** The process of training and tuning the model.
- **Parameters:** Are learned automatically by the model.
    - Examples include weights and biases.

- **Hyperparameters:**
    - Control the behavior of the learning algorithm.
    - Impact the speed and quality of the learning process.
    - Examples include batch size and learning rate.
        - **Batch size:** The number of training examples the model processes at one time before updating its weights.
- **Epochs:** One complete pass of the entire training dataset through the algorithm.

Evaluating the Model

- **Prediction Outcomes:**
    - **True Positive:** Correctly predicted as true.
    - **True Negative:** Correctly predicted as false.
    - **False Positive:** Incorrectly predicted as true (was supposed to be false).
    - **False Negative:** Incorrectly predicted as false (was supposed to be true).
- **Metrics for Classification Models:**
    - **Accuracy and Precision**
    - **Precision:** The proportion of predicted positives that were actually positive.
    - **Recall:** Measures the proportion of actual positives that were correctly identified by the model.
    - **F1 Score:** The harmonic mean of precision and recall.
    - **Confusion Matrix and AUC-ROC:** The AUC-ROC is essentially a graphed confusion matrix.
- **Metrics for Regression Models:**
    - Mean Absolute Error, Mean Squared Error, Root Mean Squared Error.
    - **R^2**: A metric indicating how well the model predicts outcomes using the input variables.

Deploying the Model

- Deploy the model to a production environment.
- **MLOps:** A set of practices for managing the ML lifecycle.
    - Focuses on automation, standardization, consistency, and reliability.
    - Enables continuous deployment of models.
    - If you change hyperparameters, it automatically tests the model.
- After deployment, you must monitor, collect data, and evaluate the model's performance.

AWS Managed AI/ML Services and Applications

- **Amazon Rekognition:** Tracks people, analyzes faces and facial emotions, and performs object detection. It can be customized.

- **Amazon Textract:** Extracts text from images, grids, tables, and forms. Offers both synchronous and asynchronous processing.
- **Amazon Comprehend:** An NLP service to find insights, useful for spam detection and sentiment analysis. Performs tokenization and Part-of-Speech (POS) tagging and can be customized. Offers both synchronous and asynchronous processing.
    - **Comprehend Medical:** A specialized version for medical text.
- **Amazon Translate:** Maintains the original tone and fluency of the text. Offers both synchronous and asynchronous processing. It's important to double-check the output, as some languages have grammatical gender or different punctuation rules.
- **Amazon Polly:** A text-to-speech service. Allows customization of language, voice, and pronunciation (pauses, emphasis, lexicons for abbreviations).
- **Amazon Transcribe:** A speech-to-text service. Powered by Automatic Speech Recognition (ASR). Can identify languages, remove personal information, and use custom vocabularies.
    - **Amazon Transcribe Medical:** A specialized version for medical speech.
- **Amazon Lex:** Uses speech-to-text to build chatbots.
    - **Intent:** The specific action the user is trying to achieve.
    - **Slots:** Variables that capture specific details from the user's input.
    - **Integrations:** Can integrate with Lambda to prompt users for more information, Amazon Connect for call center bots, and Amazon Comprehend.
- **Amazon Forecast:** Predicts future trends like sales, inventory levels, and demand.
- **Amazon Kendra:** An intelligent search service that uses NLP.
- **Amazon Personalize:** Provides advertisement personalization based on user behavior like clicks, purchases, and ratings. Uses "recipes" to suggest popular or related items.
- **Amazon SageMaker:** A platform to build ML models from scratch, covering preparing, building, training, tuning, and deploying. Supports version control and all major learning types (Supervised, Unsupervised, RL, DL).
    - **Studio:** Offers an IDE, Notebooks, Canvas, data preparation/visualization tools, and collaboration features.
    - **SageMaker Pipelines:** Automates the ML workflow.
    - **AutoML (Autopilot):** Automates model selection and hyperparameter tuning. You upload data and specify your target variable.
    - **Data Wrangler:** Simplifies the process of preparing and transforming data. It works on various data forms and allows you to import data, clean data (missing values, duplicates, outliers), perform feature engineering, and visualize data.
    - **Feature Store:** A place to store, manage, and share features. You can save features and publish or directly manage them.
    - **Endpoints:**
        - **Real-Time (Synchronous):** For chatbots, fraud detection, and recommendation systems.
        - **Asynchronous Inference:** For large files, complex models, and image or video processing.

- **Batch Transform (Batch Inference):** Asynchronous processing for large volumes of data.
- **Serverless:** Automatically scales to accommodate varying loads, ideal for sporadic traffic, minimizing cost, and simplifying deployment.

Practice Questions: AI/ML Fundamentals

Question 1 A company is developing a recommendation system for an online shopping platform. The goal is for the system to enhance suggestions based on user behavior and feedback over time. Which AI Learning strategy enables this adaptive improvement, and why?

- A. Unsupervised learning to identify patterns in user preferences
- B. Supervised learning with a fixed dataset of past purchases and ratings
- C. Supervised learning with an evolving dataset of customer interactions
- D. Reinforcement learning with rewards based on customer engagement metrics **Correct Answer:** D
- **Explanation:** Reinforcement learning is a perfect fit for adaptive recommendations. The system tries different suggestions, observes user responses (clicks, purchases), and receives "rewards" based on that engagement. Over time, it optimizes its actions to maximize user satisfaction and business goals. Unsupervised learning just finds patterns and doesn't adapt dynamically. Supervised learning with a fixed dataset doesn't adapt unless it is retrained.

Question 2 A company has developed a classification model to identify whether emails are spam or not. After training and validating the model, the company wants to evaluate its performance using a metric that represents the overall proportion of correct predictions. Which evaluation metric should the company use and why?

- A. R squared
- B. Root Mean Squared Error
- C. F1 Score
- D. Accuracy **Correct Answer:** D
- **Explanation:** Accuracy directly measures the proportion of correctly predicted labels out of all predictions. Since the company explicitly wants the overall proportion of correct predictions, accuracy is the right choice. R-squared and RMSE are used for regression problems. The F1 Score is useful when there is a class imbalance but doesn't simply measure overall correctness.

Question 3 A company is developing a sentiment analysis tool to assess customer feedback from various sources. The goal is to automatically analyze text data and derive insights about customer emotions and key themes. Which solution meets these requirements and why?

- A. Amazon Textract
- B. Amazon Comprehend
- C. Amazon SageMaker
- D. Amazon Rekognition **Correct Answer:** B
- **Explanation:** Amazon Comprehend is purpose-built for Natural Language Processing (NLP). It can perform sentiment analysis, entity recognition, and extract insights from unstructured text, which is exactly what is needed for analyzing customer feedback. Textract extracts text from documents, SageMaker is for building custom models, and Rekognition is for image and video analysis.

Question 4 Which Amazon SageMaker service is designed to streamline the data preparation process by allowing users to visually clean and transform data for machine learning and why?

- A. Amazon SageMaker Feature Store
- B. Amazon SageMaker Clarify
- C. Amazon SageMaker Pipelines
- D. Amazon SageMaker Data Wrangler **Correct Answer:** D
- **Explanation:** Amazon SageMaker Data Wrangler is specifically designed to simplify data preparation. It provides a visual interface for users to clean, transform, and explore data without writing extensive code, streamlining preprocessing before model training. Feature Store is for managing features, Clarify is for bias detection, and Pipelines is for automating workflows.

Fundamentals of Generative AI

Core Concepts

- **Token:** The smallest unit of text for a model (tokenization is the process of breaking text into tokens).
- **Context Window:** The number of tokens a Large Language Model (LLM) can take in as input.
- **Chunking:** Breaking larger datasets into smaller pieces to help manage them.
- **Embeddings:** Numerical representations of values or objects like text, images, and audio. They are useful for recommendation systems.
- **Vectors:** The numeric version of embeddings that indicate where a value or object is located in a high-dimensional space.

Model Architectures and Types

- **Transformer-Based LLM:** Processes input data using neural networks to generate a human-understandable output.
- **Foundation Models:** Large-scale pretrained models that can be adapted for various tasks.
- **Multi-modal Models:** Can handle multiple types of data such as images, text, and audio.
- **Diffusion Models:**
    - **Forward Process:** Adds noise to an image.
    - **Reverse Process:** Removes noise from an image to generate a new one.

Key Processes

- **Prompt Engineering:** The process of designing effective inputs (prompts) to get desired outputs from a model.
- **Foundation Model Lifecycle:**
    1. Data Selection
    2. Model Selection
    3. Pre-training
    4. Fine-tuning
    5. Evaluation
    6. Deployment
    7. Feedback

Use Cases, Pros, and Cons

- **Use Cases:** Visual media creation, code generation, customer interaction.
- **Advantages:** Adaptability, Responsiveness, Simplicity.
- **Disadvantages:** Hallucinations, Interpretability, Nondeterminism, Bias.

Considerations

- **Model Selection Factors:** Compliance, Constraints, Capabilities, Performance requirements.
- **Business Value and Metrics:** Efficiency and Accuracy, Conversion rate and value, Cross-domain performance, Customer lifetime value.

Generative AI in AWS

- **Amazon SageMaker Jumpstart:** Provides built-in algorithms, pre-trained models, and customized solutions. It is integrated with Amazon S3.
- **Amazon Bedrock:** Allows you to use foundation models from multiple providers.

    - **PartyRock:** A tool for testing AI models before deployment.
- **Amazon Q:** Allows users to query data using NLP and is integrated with Amazon QuickSight.
    - **Amazon Q Business:** Provides dashboard generation, summaries, and data stories.
    - **Amazon Q Developer:** Provides code generation, automation, and integration capabilities.

Benefits and Tradeoffs

- **Advantages of AWS Gen AI Services:** Accessibility, Efficiency, Cost-effectiveness, Speed to market.
- **Benefits of AWS Infrastructure:** Security, Safety, Compliance.
- **Tradeoffs:**
    - **Responsiveness and Availability:** Faster models are more expensive.
    - **Redundancy and Performance:** S3 storage for redundancy adds cost; faster Bedrock models require more expensive infrastructure.
    - **Provisioned Throughput and Custom Models:** Reserving computational resources costs money; custom models may have data prep and fine-tuning costs.

Practice Questions: Generative AI

Question 1 You are developing a chatbot that will generate responses and provide summaries for customer queries. What foundation model lifecycle stage is necessary for the chatbot to improve over time by learning from customer interactions?

- A. Pre-training
- B. Model Selection
- C. Fine-tuning
- D. Evaluation **Correct Answer:** C
- **Explanation:** Fine-tuning is when a pre-trained model is further trained on domain-specific data, like customer conversations. This allows the chatbot to learn from customer interactions over time and improve its performance for the specific use case. Pre-training is the initial stage of general data, model selection is about choosing the starting model, and evaluation is for testing performance.

Question 2 Which term refers to the process of converting words or phrases into numerical vectors, allowing models to understand the meaning of text in a high-dimensional space?

- A. Chunking

- B. Embedding
- C. Tokenizing
- D. Diffusion **Correct Answer:** B
- **Explanation:** Embedding is the process of mapping words or phrases into dense numerical vectors. These vectors capture semantic meaning. Chunking is breaking text into larger pieces, tokenizing is splitting text into smaller units, and diffusion is a type of generative model.

Question 3 A company is using a generative AI model to summarize legal documents, but sometimes the model produces inaccurate or completely fabricated information. What is the most likely cause of this issue?

- A. Hallucination
- B. Nondeterminism
- C. Bias
- D. Interpretability **Correct Answer:** A
- **Explanation:** In generative AI, hallucination refers to the model generating plausible-sounding but factually incorrect or fabricated information. This is the most likely cause for inaccurate content in legal summaries. Nondeterminism refers to variability in outputs, bias refers to skewed patterns from training data, and interpretability refers to how easily humans can understand a model's decision.

Question 4 Which AWS service is best for developers looking to quickly start using pre-trained models for generative AI tasks like text generation or image creation without needing extensive machine learning expertise?

- A. Amazon Bedrock
- B. Amazon Q
- C. Amazon EC2
- D. Amazon SageMaker Jumpstart

Question 5 A business is deciding between different AWS generative AI services to build a custom model that requires high availability across multiple regions. Which cost tradeoff should the company be most concerned about?

- A. Responsiveness
- B. Token-Based pricing
- C. Regional coverage
- D. Regional throughput **Correct Answer:** C
- **Explanation:** Deploying a model across multiple AWS regions increases infrastructure and replication costs. The business must pay extra to maintain high availability in more than one region, making this the main cost tradeoff for multi-region resilience. Responsiveness is about latency, not a direct multi-region cost tradeoff.

Guidelines for Responsible AI

Core Principles

- **Responsible AI Systems** should be Ethical, Transparent, and Trustworthy.
- **Fairness and Bias Mitigation:**
    - Ensure inclusivity and diversity in data.
    - Strive for balanced representation, use high-quality sources, and perform ethical labeling.
- **Explainability:** The ability to explain how a model arrived at a decision.
    - Making a model more transparent makes it more explainable but may affect performance.
- **Robustness and Veracity:** AI should adapt to challenging conditions and be reliable and truthful.
- **Controllability:**
    - Involves model selection and environmental selection.

Understanding Bias and Variance

- **Bias:** The difference between the predictive and actual values (level of error).
    - High Bias is also known as Underfitting.
    - To reduce high bias, you can increase the number of features or use a more complex model.
- **Variance:** The extent to which a model's predictions change when trained on different data.
    - High Variance means the model fits the training data well but performs poorly on new, unseen data (Overfitting).
    - To reduce high variance, you can select fewer (only important) features, get more data, or increase the dataset size with data augmentation.
- The goal is to minimize both bias and variance.
- **Subgroup analysis** involves analyzing data and model performance across different subgroups.

Types of Bias

- **Measurement Bias:** Caused by inaccurate measurement during data collection.
- **Sampling Bias:** Occurs when the training data is not representative of the whole population.
- **Confirmation Bias:** Focusing on data that confirms your beliefs while ignoring contradictory data.

- **Observer Bias:** When the person collecting data introduces bias.

Risks in Generative AI

- **Hallucinations:** The model generates plausible but factually incorrect information.
- **Prompt Leaking:** The model inadvertently reveals too much context or history about previous interactions.
- **Model Exposure:** Exposure of confidential information by the model.
- **Intellectual Property:** The model generates material that resembles copyrighted works.

Building Responsible AI with AWS Tools

- **Amazon SageMaker Clarify:**
  - Detects bias in data and in model predictions.
  - Can help detect which feature is causing the problem.
  - Shows a bar chart of which features are impacting a decision most.
- **Amazon SageMaker Ground Truth:**
  - Simplifies and speeds up the process of labeling data.
  - Uses human labelers and can provide machine assistance with suggestions.
  - Offers automated labeling capabilities.
  - **Ground Truth Plus:** Provides access to expert labelers.
- **Amazon Augmented AI (A2I):**
  - Triggers human review of model predictions, either randomly or based on confidence scores.
- **Amazon Bedrock Guardrails:**
  - Can block against prompt attacks and unwanted outputs.

AI Governance

- **SageMaker Model Cards:**
  - Helps document a model's purpose, risk ratings, limitations, ethical considerations, and performance metrics.
- **SageMaker Model Monitor:**
  - Compares live production data to the training data.
  - Sends alerts if data drift occurs (when production data is completely different) or if bias is detected.

Practice Questions: Responsible AI

Question 1 A bank's machine learning model predicts auto loan approvals. The model rejects rural applicants more often than urban ones, despite similar credit histories and incomes. Which type of bias is affecting the model output?

- A. Measurement bias
- B. Confirmation bias
- C. Observer bias
- D. Sampling bias **Correct Answer:** D
- **Explanation:** Sampling bias occurs when the training data is not representative of the population. The fact that the model rejects rural applicants more often despite similar financial profiles suggests the training data may have underrepresented them or contained historical biases. This causes the model to learn unfair patterns. There is no mention of measurement errors or human biases like confirmation or observer bias.

Question 2 A company is developing a solution to classify medical images for detecting skin cancer. The solution must ensure high accuracy in labeling and minimize the risk of incorrect classifications, especially for rare types of cancer. Which solution will meet these requirements?

- A. Image analysis by using Amazon Rekognition
- B. Text analysis by using Amazon Comprehend Medical
- C. Data augmentation by using an Amazon Bedrock knowledge base
- D. Human-in-the-loop validation by using Amazon SageMaker Ground Truth Plus **Correct Answer:** D
- **Explanation:** For critical medical images and rare conditions, high accuracy is essential. Amazon SageMaker Ground Truth Plus enables human-in-the-loop validation, allowing domain experts to review predictions or label data. This is crucial for rare conditions where automated models may struggle and helps minimize misclassifications. Rekognition is for general-purpose image analysis, not medical imaging. Comprehend Medical works with text, not images.

Applications and Operations of Foundation Models

Design Considerations

- **Modality:** The type of data a model is trained to handle (e.g., text, image).
- **Latency:** The time it takes for a model to produce an output.
- **Multi-lingual:** The model's ability to handle multiple languages.
- **Model size and complexity**
- **Customization:** The ability to adapt the model to specific tasks.

- **Input/Output Length:** The amount of data the model can process and generate.

Inference Parameters

- **Temperature:** Controls randomness. Lower is more predictable, higher is more creative.
- **Top-k:** Narrows the selection to the 'k' most likely tokens for the output.
- **Top-p:** Narrows the selection to a percentage of the most likely candidates for the next token.

Architectural Patterns

- **Retrieval Augmented Generation (RAG):** Enhances model outputs by retrieving information from a knowledge base.
- **Vector Databases and Embeddings:**
    - Embeddings capture semantic meanings.
    - Vector databases store these embeddings for efficient search.
    - AWS Options: Amazon OpenSearch, Amazon Aurora, Amazon Neptune, Amazon DocumentDB, Amazon RDS for PostgreSQL.

Customization Methods

- **Pre-training:** High cost, full control over model behavior.
- **Fine-tuning:** Moderate cost, balances control and efficiency.
- **In-context learning (Prompting):** Low cost.
- **RAG:** Cost-effective way to add knowledge.

Agents for Amazon Bedrock

- Agents can perform tasks like RAG, code generation, and other multi-step automated actions.

Prompt Engineering

Techniques

- **Key Components:** Instructions, Context, Negative Prompts (what to avoid).
- **Zero-Shot Prompting:** Providing no examples in the prompt.
- **Single-Shot Prompting:** Providing one example.
- **Few-Shot Prompting:** Providing more than one example.
- **Chain-of-Thought Prompting:** Asking the model to explain its thought process step-by-step.

- **Prompt Templates:** Reusable prompt structures.

Improving Outputs

- Use more specific and concise prompts.
- Experiment with different prompts.
- Use guardrails (negative prompts).
- Use multiple comments for related queries.

Risks

- **Exposure:** Confidential data is leaked.
- **Poisoning:** Malicious data is inserted during training.
- **Hijacking/Jailbreaking:** Manipulating a prompt to generate a desired (and often unsafe) output or bypass safety measures.

Training, Fine-tuning, and Evaluation

Training and Fine-tuning Methods

- **Pre-training:** General learning from massive datasets.
- **Fine-tuning:** Customizing the model for a specific task.
- **Continuous Pre-Training:** Keeping the model up-to-date with new data.
- **Instruction Tuning:** Teaching a model to follow instructions better.
- **Domain Adaptation:** Adapting models for specific domains (e.g., medical, legal).
- **Transfer Learning:** Fine-tuning an already fine-tuned model for other related topics.
- **Reinforcement Learning from Human Feedback (RLHF):** The model adapts based on human preferences.

Data Preparation for Fine-tuning

- **Data Curation:** Filtering out unrelated context.
- **Data Governance**
- **Data Size and Representativeness**
- **Data Labeling**

Evaluation Methods

- **Human Evaluation:** Judges how well responses align with real-world expectations; can be slow and subjective.
- **Benchmark Datasets:** Prebuilt datasets that are objective.

- **Performance Metrics:**
    - **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** Tests for recall ability by measuring overlaps between generated and reference texts; great for summarization.
    - **BLEU (Bilingual Evaluation Understudy):** Focused on the quality of text, commonly used for translation.
    - **BERTScore:** Uses pre-trained BERT models to compare semantic similarities.

Business Objective Alignment

- **Productivity:** How efficient the model is at performing tasks.
- **User Engagement:** How often users interact with the model.
- **Task Engineering:** How effectively the model can perform business-related tasks.

Practice Questions: Foundation Models

Question 1 Which factors should you prioritize when selecting a pre-trained foundation model for real-time language translation?

- A. Model size and complexity
- B. Latency and multi-lingual capabilities
- C. Input/output length
- D. RAG compatibility **Correct Answer:** B
- **Explanation:** For real-time translation, low latency (fast response) is critical. The model must also have strong multilingual coverage to handle diverse languages accurately. While model size is a factor, latency is more important for real-time needs. RAG is not relevant for direct translation.

Question 2 Which prompt engineering technique involves breaking down complex problems into sequential reasoning steps to improve the model's response accuracy?

- A. Prompt templates
- B. Few-shot prompting
- C. Chain-of-thought prompting
- D. Negative prompting **Correct Answer:** C
- **Explanation:** Chain-of-thought prompting guides the model to reason step-by-step, which improves accuracy on tasks requiring logic or multi-step reasoning. Prompt templates are for consistency, few-shot prompting provides examples, and negative prompting specifies what to avoid.

Question 3 Which fine-tuning method involves customizing a model for a specific field, such as legal or medical applications?

- A. Transfer learning
- B. Instruction tuning
- C. Domain adaptation
- D. Continuous pre-training **Correct Answer:** C
- **Explanation:** Domain adaptation is a fine-tuning method that specializes a model for a specific field or domain (e.g., legal, medical) so it performs better with domain-specific terminology and context. Transfer learning is a broader concept, instruction tuning focuses on following instructions, and continuous pre-training is about updating the model with new data.

Question 4 A company needs a scalable, objective method to evaluate the performance of text generation models based on semantic similarity and linguistic quality. Which of the following approaches is the best choice for automated evaluation?

- A. ROUGE
- B. BLEU
- C. BERTScore
- D. Human Evaluation **Correct Answer:** C
- **Explanation:** BERTScore uses contextual embeddings to measure semantic similarity between generated and reference text. It captures meaning better than surface-level metrics, making it well-suited for scalable, automated evaluation. ROUGE and BLEU focus more on word overlap. Human evaluation is not scalable or automated.

Securing AI Systems

- **Identity and Access Management (IAM):** Involves managing identities and their access rights to ensure appropriate level of permissions to resources is granted
    - **Identity-based:** JSON permissions policy documents
    - **Fine-grained (Principle of Least Privilege):** Granular control over access to resources
    - **AI Resources**
    - **Role-based Access Control:** Scales well
    - **Attribute-based Access Control:** Analyzes set of attributes and enables access based on logic, can leverage tags (production tags to certain resources)
    - **Temporary Access**

- **Key Management Service (KMS):** Create, Store, and Manage keys (Encryption)

    - **Customer Managed Key:** Can Manage Keys and View Metadata
    - **AWS Managed Key:** Can Manage key but can't view metadata
    - **AWS Owned Key:** Can't Manage key or view metadata
- **CloudHSM:** Performs cryptographic operations, securely stores and manages cryptographic keys

- **Macie:** Used in S3 and uses ML to classify sensitive data
- **AWS PrivateLink:** Allows private network connection between VPCs and AWS Services

- **Data Origins:** How data is collected, cleaned, and transformed
- **Data Source Citation:** Acknowledge source of data
- **Data Lineage:** Tracking everything about the data
- **Data Cataloging:** Organizes and documents AI data components
    - **SageMaker Model Cards:** Can do all above

Data Access Control

- IAM
- Encryption
- Network Controls
- Monitor

Compliance

- **Personally Identifiable Information (PII):** Similar information is Protected Health Information (PHI)
- **National Institute of Standards and Technology (NIST):** Outlines security controls and guidelines for privacy risk management and protection
- **Health Insurance Portability and Accountability Act (HIPAA)**
- **General Data Protection Regulation (GDPR):** PII for European Union
- **Privacy Reference Architecture:** Guide and best practices for designing an AWS privacy controls-centric architecture

Threat Detection

- **Training Data Poisoning:** False or malicious data is injected into AI Models
- **Misuse:** Manipulating AI System
- **Misconfiguration**
- **Amazon GuardDuty:** Detects anomaly
- **Amazon Inspector:** Vulnerability management
    - **Identify, Classify, Remediate, Mitigate**
- **Amazon Detective:** Incident response

Incident Response

1. **Preparation:** Understand the threats, set up defenses
2. **Detection and Analysis:** Monitor for activity, analyze information
3. **Containment, Eradication, and Recovery:** Limit the spread, remove malware, stop attack
4. **Post-incident:** Retain evidence, reflect

OWASP Top 10 for LLM Applications

1. **Prompt Injection:** Manipulates LLM through inputs,
2. **Insecure Output Handling:** LLM output is accepted exposing backend systems
3. **Training Data Poisoning**
4. **Model Denial of Service**
5. **Supply Chain Vulnerabilities**
6. **Sensitive Information Disclosure**
7. **Insecure Plugin Design**
8. **Excessive Agency**
9. **Over-reliance**
10. **Model Theft**

- **Amazon Bedrock Guardrails:** Has input sanitization and validation, Fine-grained access controls, rate limiting, output scanning

Application Security

- **Data**
    - **Data preparation**
- **Build and Test**
    - **Model Build, Model Evaluation, Model Selection**
- **Deployment**
- **Monitoring and Response**
- Securing data used to train model, code of the model, and backups of model
- **AWS Web Application Firewall (WAF):** Provides defense against common web app attacks by using Web ACL for custom rules
- **AWS Shield:** Protects against DDoS
    - DDoS Response team for advanced tier
    - Integrated Protection
- **Amazon Cognito:**
    - IAM for Mobile and Web Apps
    - Adaptive Protections

Securing AI Infrastructure

- **Encryption** (Like KMS)
- **Access Control** (Like IAM)
- **Edge** (Network Segmentation)

Governance and Compliance for AI Systems

AWS Config: Helps facilitate data governance by providing detailed information relating to configuration (Governance)
Amazon Inspector: Vulnerability Management with Scanning, Assessments, Findings (Risk)
Compliance: Payment Card Industry Data Security Standard (PCI-DSS), National Institute of Standards and Technology (NIST), Health Information Portability and Accountability Act (HIPAA), General Data Protection Regulation (GDPR)

International Organization for Standardization (ISO)

- **ISO/IEC 27001**
    - Specifics requirements, establishing, implementing, maintaining, continual improvement
- **ISO/IEC 27002**
    - Supporting standard, guidelines, best practices

System and Organization Controls (SOC)

- **SOC 1:** Financial Reporting
    - Type 1: Point in time snapshot
    - Type 2: Time Range
- **SOC 2:** Trust Services
    - Type 1: Point in time snapshot
    - Type 2: Time Range
- **SOC 3:** Public Summary

AWS Artifact: Provides on demand access to security and compliance documents, certifications and reports

Accountability and Transparency

- **Accountability:** Logging, Organization, Legal Ramification
- **Transparency:** Terms of Service, Removal and Retention, Sharing Customer Data

Algorithm Accountability Laws

- **European Union's proposed Artificial Intelligence Act:** First comprehensive legal framework for AI, addressing risks
- **NYC Automated Decision Systems Law:** Aims to prevent biases in AI for employment decisions

AWS Services for GRC

- **AWS Config:** Manage and view AWS resources configurations
- **Amazon Inspector:** Security Scanning, Assessments, Findings
- **Amazon Detective:** Aids in investigation of security events and helps identify root cause by using logs
- **Amazon Audit Manager:** Audit usage for GRC, gathers evidence for auditing
- **AWS Artifact:** Provides on demand access to security and compliance documents, reports, and certifications for auditors or regulators
- **Trusted Advisor:** Evaluations environment for best practices, Remediates
- **AWS CloudTrail:** Records all API calls by users and AWS Services
- **GuardDuty:** Monitoring and alerting threats
- **AWS Security Hub:** Centralized dashboard for all of these services

Data Lifecycles

1. Collect
2. Process
3. Store
4. Consumption
5. Archive
6. Disposal

Logging and Monitoring

- Input and output behavior
- Performance metrics
- Security events
- Infrastructure
- Responsible AI usage

Governance Protocols

- Procedure -> Process -> Policy
- Review Cadence
    - Many Aspects to consider
    - Schedule for reviewing policies
    - Many representations

- Review Strategies
    - Technical reviews: Model performance, quality of data and algorithms
    - Non-technical reviews: Helps ensure solution aligns with regulatory requirements and organizational policies
    - Testing: Validates solutions' outputs
- Review Intervention

Governance Frameworks

- **Generative AI Security Scoping Matrix**
    - **Application:** Public Consumer, Enterprise
    - **Model:** Pre-trained, Fine-tuned, Self-trained

Transparency Standards

- Document, Feedback, Publishing information

Team Training

Bias

- **Algorithmic:** Systemic and unfair discrimination in algorithm's outcomes
- **Confirmation:** Cognitive bias based on training data that reinforces stereotypes
- **Selection:** Occurs when training data selected is skewed and does not adequately represent the broader population

---

Practice Questions: Security, Compliance and Governance

Question 1 A healthcare organization is developing an AI-based solution to assist in patient diagnostics. The solution will access sensitive healthcare data, and the organization needs to ensure compliance with data protection regulations. Which regulation should the organization comply with to protect patient data in the United States?

- A.General Data Protection Regulation (GDPR)
- B. Health Insurance Portability and Accountability Act (HIPAA)
- C. Payment Card Industry Security Standard (PCI-DSS)
- D. National Institute of Standards and Technology (NIST) **Correct Answer:** B

Explanation: HIPAA governs the protection of **protected health information (PHI)** in the United States. Any healthcare organization handling patient data must comply with HIPAA to ensure confidentiality, integrity, and security of health records. GDPR applies in the EU, PCI-DSS is for payment card data, and NIST provides security guidelines but is not a regulation.

Question 2 A company is planning to deploy a generative AI application on AWS to handle customer service requests. The solution will access and process large amounts of customer data, and the company needs to ensure that only authorized users and services can access the application. Which of the following AWS services should the company use to manage user access to the generative AI application?

- A. AWS Identity and Access Management (IAM)
- B. AWS Key Management Service (KMS)
- C. Amazon Elastic Load Balancer (ELB)
- D. AWS Systems Manager **Correct Answer:** A

Explanation: AWS Identity and Access Management (IAM) is the service that allows organizations to securely manage access to AWS resources. It provides fine-grained control over who can access specific services and resources, ensuring that only authorized users and applications can interact with the generative AI solution. KMS is designed for encryption key management, ELB is used for traffic distribution, and Systems Manager is focused on operational management and automation, making IAM the correct choice for access control.

Question 3 A company is deploying a large language model (LLM) for customer support. The developers are concerned about the risk of prompt injection. They need a way to enforce input validation and control the prompts before they are sent to the model. Which AWS service can help mitigate the risk of prompt injection by enforcing input validation and ensuring only approved prompt templates are used?

- A. AWS WAF (Web Application Firewall)
- B. AWS Key Management Service (KMS)
- C. Amazon Bedrock Guardrails
- D. AWS S3 **Correct Answer:** C

Explanation: Amazon Bedrock Guardrails provides a way to enforce policies, validate inputs, and control the prompts sent to large language models. It helps prevent risks such as prompt injection by ensuring that only approved prompt templates and safe inputs are passed to the model. AWS WAF is designed for filtering and blocking malicious web traffic, KMS is used for encryption key management, and S3 is for object storage, none of which address prompt control for LLMs.

Question 4 A global company is developing an AI-based system that will process sensitive customer data across multiple regions. The legal and compliance teams need to ensure that the system complies with data protection laws and regulations. Which two considerations should the company prioritize in order to meet legal and regulatory requirements? (Select two)

- A. Compliance with data privacy regulations
- B. Data sovereignty requirements
- C. Optimizing AI model performance

- D. Increasing cloud storage capacity for data processing **Correct Answers:** A and B

Explanation: To meet legal and regulatory requirements, the company must ensure compliance with data privacy regulations such as GDPR, HIPAA, or other region-specific laws that govern how sensitive customer data is handled. Additionally, data sovereignty requirements must be considered, as many countries mandate that customer data be stored and processed within their borders. While model performance optimization and cloud storage capacity are important for scalability and efficiency, they do not directly address compliance or legal obligations.

Question 5 A company is deploying a large language model (LLM) for customer support and needs to ensure full accountability and traceability of all API calls made by users and AWS services interacting with the model. They need to record both successful and failed API calls and ensure all service usage is logged for auditing and compliance purposes.Which AWS service can help the company achieve this by logging all API activity across AWS services, including LLM interactions?

- A. Amazon CloudWatch
- B. AWS Config
- C. AWS CloudTrail
- D. AWS Lambda **Correct Answer:** C

Explanation: AWS CloudTrail is the service that records all API calls made within an AWS environment, including those initiated by users, roles, or other AWS services. It captures both successful and failed requests, providing detailed logs that are essential for auditing, compliance, and security investigations. Amazon CloudWatch is focused on monitoring performance metrics and logs, AWS Config tracks configuration changes to resources, and AWS Lambda is a compute service for running code, none of which provide comprehensive API activity logging across services.

Question 6 A company is implementing an information security management system (ISMS) to meet the requirements of ISO 27001 and ISO 27002. The company wants to understand the differences between the two standards in order to effectively apply them to their security practices. Which of the following best describes the differences between ISO 27001 and ISO 27002?

- A. ISO 27001 is a set of recommended practices for security controls, while ISO 27002 outlines the requirements for risk assessment
- B. ISO 27001 specifies the requirements for establishing, implementing, and maintaining an ISMS, while ISO 27002 provides guidelines for implementing security controls
- C. ISO 27001 focuses on payment card transactions, while ISO 27002 is a more detailed guide on how to deploy secure containers
- D. ISO 27001 is a code of practice for information security, while ISO 27002 is a management framework for ISMS **Correct Answer:** B

Explanation: ISO 27001 defines the formal requirements for establishing, implementing, maintaining, and continually improving an information security management system (ISMS). It is

the certifiable standard that organizations must comply with. ISO 27002, on the other hand, provides best-practice guidelines and recommendations for selecting and implementing information security controls. In other words, ISO 27001 tells organizations what they need to do to achieve certification, while ISO 27002 helps them understand how to implement the required controls effectively.

Question 7 A Finance Technology organization is utilizing a Large Language Model (LLM) for generating content. What critical OWASP LLM vulnerability should the SecOps team be most concerned about when defending against potential malicious input and system manipulations?

- A. Insecure Output Handling
- B. Training Data Poisoning
- C. Prompt Injection
- D. Sensitive Information Disclosure **Correct Answer:** C

Explanation: Prompt injection is a critical OWASP LLM vulnerability where attackers manipulate the input prompts to override instructions, execute unintended actions, or exfiltrate sensitive data. Since LLMs rely heavily on natural language inputs, malicious prompts can exploit this trust to compromise the system. While insecure output handling, training data poisoning, and sensitive information disclosure are also important risks, prompt injection represents the most immediate and critical concern for defending against malicious input and manipulations in LLM-based systems.

Question 8 Your team has been tasked with developing a generative AI solution that will be built using self-trained model data. As part of this process, the organization needs to ensure that the model adheres to AI algorithm accountability laws and regulations, ensuring transparency, fairness, and compliance. Which of the following laws and regulations should your organization consider to ensure compliance with AI algorithm accountability during the development of the self-trained AI model? (Select two)

- A. New York City's Automated Decision Systems Law
- B. Florida's AI Command Law
- C.Health Insurance Portability and Accountability Law
- D.European Union's Artificial Intelligence Act **Correct Answers:** A and D

Explanation: New York City's Automated Decision Systems Law requires transparency and accountability for automated systems that impact decision-making, ensuring fairness and reducing bias. The European Union's Artificial Intelligence Act is one of the most comprehensive regulatory frameworks for AI, focusing on risk management, transparency, and accountability in AI systems. Florida's AI Command Law does not exist, and HIPAA is a healthcare data privacy regulation, not an AI accountability framework. Therefore, the two relevant laws for AI algorithm accountability are New York City's Automated Decision Systems Law and the EU AI Act.

Question 9 A start-up company has governance demands, and the organization must select AWS cloud services that provide robust control and visibility into their object storage

infrastructure. The solution must enable precise retention periods and comprehensive configuration tracking. Which AWS services are best suited to meet these demands? (Select two)

- A. Amazon Inspector
- B. AWS Config
- C. Amazon S3 Object Locking
- D. AWS IAM **Correct Answers:** B and C

Explanation: AWS Config provides continuous monitoring and recording of configuration changes across AWS resources, enabling comprehensive tracking and auditing of the object storage infrastructure. Amazon S3 Object Locking allows organizations to enforce retention periods and prevent objects from being deleted or modified, supporting governance and compliance requirements. Amazon Inspector focuses on security assessment, and AWS IAM manages access control but does not provide retention enforcement or detailed configuration tracking for storage.

Question 10 A start-up is focused on building its own generative AI models and seeks to ensure proper security and governance during development. The company is advised to use the GenAI Security Scoping Matrix to guide their approach. Based on the GenAI Security Scoping Matrix, which of the following scopes should the company prioritize to address security and compliance when developing their own generative AI models? (Select three)

- A. Scope 1: Public Consumer App
- B. Scope 2: Enterprise App
- C. Scope 3: Pre-trained Model
- D. Scope 4: Fine-tuned Model
- E. Scope 5: Self-trained Model **Correct Answers:** C and D and E

Explanation: When developing generative AI models, the primary focus should be on the security and governance of the models themselves. Scope 3 (Pre-trained Model) involves managing risks associated with using external pre-trained models. Scope 4 (Fine-tuned Model) addresses security considerations during the fine-tuning process. Scope 5 (Self-trained Model) covers full lifecycle governance and compliance for models trained from scratch. Scopes 1 and 2 focus on application deployment rather than model development, so they are less critical for securing and governing the AI models themselves.

Question 11 In the context of AI systems, data logging involves the systematic recording of data related to the processing of an AI workload. Which two factors should be prioritized when setting up data logging for AI workloads? (Select two)

- A. Tracking outputs
- B. User preference configuration
- C. Data backup schedule
- D. Model performance metrics **Correct Answers:** A and D

Explanation: When setting up data logging for AI workloads, it is essential to track the outputs of the AI system to ensure accountability, traceability, and reproducibility of results. Additionally, logging model performance metrics allows teams to monitor accuracy, detect anomalies, and improve the AI system over time. User preference configuration and data backup schedules are important for usability and disaster recovery but are not central to AI data logging for monitoring and accountability purposes.