Exam Guide (AIF-C01)

AWS Certified AI Practitioner

AWS Certified AI Practitioner: Exam Guide (AIF-C01)

Table of Contents

AWS Certified AI Practitioner (AIF-C01)

The AWS Certified AI Practitioner (AIF-C01) exam is designed for individuals who want to demonstrate a foundational understanding of AI concepts and AWS AI tools. This certification focuses on practical business applications of AI.

Topics

Introduction

The AWS Certified AI Practitioner (AIF-C01) exam is designed for individuals who want to demonstrate a foundational understanding of AI concepts and AWS AI tools. This certification focuses on practical business applications of AI.

The exam also validates a candidate's ability to complete the following tasks:

- Describe AI, ML, and generative AI (GenAI) concepts, methods, and strategies in general and on AWS.
- Identify the appropriate use of AI/ML and GenAI technologies to solve business problems.
- Determine the correct types of AI/ML technologies to apply to specific use cases.

- Use AI, ML, and GenAI technologies responsibly.

Target candidate description

The target candidate should have up to 6 months of exposure to AI/ML technologies on AWS. The target candidate uses but does not necessarily build AI/ML solutions on AWS.

Recommended AWS knowledge

The target candidate should have the following AWS knowledge:

- Familiarity with the core AWS services (for example, Amazon EC2, Amazon S3, AWS Lambda, Amazon Bedrock, and Amazon SageMaker AI) and AWS core services use cases
- Familiarity with the AWS shared responsibility model for security and compliance in the AWS Cloud
- Familiarity with AWS Identity and Access Management (IAM) for securing and controlling access to AWS resources
- Familiarity with AWS service pricing models

Job tasks that are out of scope for the target candidate

The following list contains job tasks that the target candidate is not expected to be able to perform. This list is non-exhaustive. These tasks are out of scope for the exam:

- Developing or coding AI/ML models or algorithms
- Implementing data engineering or feature engineering techniques
- Performing hyperparameter tuning or model optimization
- Building and deploying AI/ML pipelines or infrastructure
- Conducting mathematical or statistical analysis of AI/ML models
- Implementing security or compliance protocols for AI/ML systems
- Developing and implementing governance frameworks and policies for AI/ML solutions

Exam content

Question types

The exam contains one or more of the following question types:

- **Multiple choice:** Has one correct response and three incorrect responses (distractors).
- **Multiple response:** Has two or more correct responses out of five or more response options. You must select all the correct responses to receive credit for the question.
- **Ordering:** Has a list of 3–5 responses to complete a specified task. You must select the correct responses and place the responses in the correct order to receive credit for the question.
- **Matching:** Has a list of responses to match with a list of 3–7 prompts. You must match all the pairs correctly to receive credit for the question.

Unanswered questions are scored as incorrect. There is no penalty for guessing. The exam includes 50 questions that affect your score.

Unscored content

The exam includes 15 unscored questions that do not affect your score. AWS collects information about performance on these unscored questions to evaluate them for future use as scored questions. The unscored questions are not identified on the exam.

Exam results

The AWS Certified AI Practitioner (AIF-C01) exam has a pass or fail designation. The exam is scored against a minimum standard established by AWS professionals who follow certification industry best practices and guidelines.

Your results for the exam are reported as a scaled score of 100–1,000. The minimum passing score is 700. Your score shows how you performed on the exam as a whole and whether you passed. Scaled scoring models help equate scores across multiple exam forms that might have slightly different difficulty levels.

Your score report could contain a table of classifications of your performance at each section level. The exam uses a compensatory scoring model, which means that you do not need to achieve a passing score in each section. You need to pass only the overall exam.

Each section of the exam has a specific weighting, so some sections have more questions than other sections have. The table of classifications contains general information that highlights your strengths and weaknesses. Use caution when you interpret section-level feedback.

Content outline

This exam guide includes weightings, content domains, tasks, and skills for the exam. This guide does not provide a comprehensive list of the content on the exam.

The exam has the following content domains and weightings:

- Content Domain 1: Fundamentals of AI and ML (20% of scored content)
- Content Domain 2: Fundamentals of GenAI (24% of scored content)
- Content Domain 3: Applications of Foundation Models (28% of scored content)
- Content Domain 4: Guidelines for Responsible AI (14% of scored content)
- Content Domain 5: Security, Compliance, and Governance for AI Solutions (14% of scored content)

Content Domain 1: Fundamentals of AI and ML

Domain 1 covers the fundamentals of AI and ML and represents 20% of the scored content on the exam.

Tasks

- Task Statement 1.1: Explain basic AI concepts and terminologies.
- Task Statement 1.2: Identify practical use cases for AI.
- Task Statement 1.3: Describe the AI/ML development lifecycle.

Task Statement 1.1: Explain basic AI concepts and terminologies.

Objectives:

- Define basic AI terms (for example, AI, ML, deep learning, neural networks, computer vision, natural language processing [NLP], model, algorithm, training and inferencing, bias, fairness, fit, large language model [LLM], generative AI [GenAI], agentic AI).

- Describe the similarities and differences between AI, ML, GenAI, deep learning, and agentic AI.
- Describe various types of inferencing (for example, batch, real-time, asynchronous, serverless).
- Describe the different types of data in AI models (for example, labeled and unlabeled, tabular, time-series, image, text, structured and unstructured).
- Describe different types of AI/ML learning (for example, supervised learning, unsupervised learning, reinforcement learning methods).

Task Statement 1.2: Identify practical use cases for AI.

Objectives:

- Recognize applications where AI/ML can provide value (for example, assist human decision making, solution scalability, automation).
- Determine when AI/ML solutions are not appropriate (for example, cost-benefit analyses, situations when a specific outcome is needed instead of a prediction).
- Select the appropriate AI/ML techniques for specific use cases (for example, regression, classification, clustering).
- Identify examples of real-world AI applications (for example, computer vision, NLP, speech recognition, recommendation systems, fraud detection, forecasting, knowledge bases, agentic AI).
- Explain the capabilities of AWS managed AI/ML services (for example, Amazon SageMaker AI, Amazon Transcribe, Amazon Translate, Amazon Comprehend, Amazon Lex, Amazon Polly).
- Identify when traditional ML models or foundation models (FMs) are appropriate for a specific use case (for example, based on regulatory concerns, explainability requirements, operational constraints).

Task Statement 1.3: Describe the AI/ML development lifecycle.

Objectives:

- Describe and differentiate components of an AI/ML pipeline.
- Describe sources of FM models (for example, open source pre-trained models, training custom models).
- Describe methods to use a model in production (for example, managed API service, self-hosted API).

- Identify relevant AWS services and features for each stage of an AI/ML pipeline (for example, Amazon Bedrock, Amazon Q, Amazon Quick, Kiro, SageMaker AI).
- Describe fundamental concepts of ML operations (MLOps) (for example, experimentation, repeatable processes, scalable systems, managing technical debt, achieving production readiness, model monitoring, model re-training).
- Describe model performance metrics (for example, accuracy, precision, recall, F1 score) and business metrics (for example, cost per user, development costs, customer feedback, return on investment [ROI]) to evaluate ML models.

Content Domain 2: Fundamentals of GenAI

Domain 2 covers the fundamentals of GenAI and represents 24% of the scored content on the exam.

Tasks

Task Statement 2.1: Explain the basic concepts of generative AI (GenAI).

Objectives:

- Define foundational GenAI concepts (for example, tokens, chunking, embeddings, vectors, prompt engineering, transformer-based large language models [LLMs], foundation models [FMs], multi-modal models, diffusion models).
- Identify potential use cases for GenAI models (for example, image, video, and audio generation; summarization; AI assistants; translation; code generation; customer service agents; search; recommendation engines).
- Describe the FM lifecycle (for example, data selection, model selection, pre-training, fine-tuning, evaluation, deployment, feedback).
- Describe the token-based pricing model and its effect on cost and performance for inference.

- Describe the role of context engineering in FM applications.
- Define foundational agentic AI concepts (for example, multi-agent system patterns for complex AI applications, Model Context Protocol [MCP] and its role in connecting agents to external systems, multi-agent communication patterns, memory management, tool usage, and workflow orchestration).

Task Statement 2.2: Understand the capabilities and limitations of GenAI for solving business problems.

Objectives:

- Describe the advantages of GenAI (for example, adaptability, responsiveness, conversational capabilities, ability to generate content).
- Identify disadvantages of GenAI solutions (for example, hallucinations, interpretability, inaccuracy, nondeterminism).
- Identify factors to consider when selecting GenAI models (for example, model types, performance requirements, capabilities, constraints, compliance, cost, latency, model complexity).
- Determine business value and metrics for GenAI applications (for example, cross-domain performance, ROI, efficiency, conversion rate, average revenue per user, accuracy, customer lifetime value).

Task Statement 2.3: Describe AWS infrastructure and technologies for building GenAI applications.

Objectives:

- Identify AWS services and features to develop GenAI applications (for example, Amazon Bedrock, Amazon SageMaker AI, SageMaker JumpStart, Amazon Quick, Kiro, Strands Agents, Amazon Bedrock AgentCore).
- Describe the advantages of using AWS GenAI services to build applications (for example, accessibility, lower barrier to entry, efficiency, cost-effectiveness, speed to market, ability to meet business objectives).
- Describe the benefits of AWS infrastructure for GenAI applications (for example, security, compliance, responsibility, safety).

- Describe cost tradeoffs of AWS GenAI services (for example, responsiveness, availability, redundancy, performance, regional coverage, token-based pricing, provision throughput, custom models).

Content Domain 3: Applications of Foundation Models

Domain 3 covers applications of foundation models and represents 28% of the scored content on the exam.

Tasks

Task Statement 3.1: Describe design considerations for applications that use foundation models (FMs).

Objectives:

- Identify selection criteria to choose FMs (for example, cost, modality, latency, multi-lingual, model size, model complexity, customization, input/output length, prompt caching).
- Describe the effect of inference parameters on model responses (for example, temperature, input/output length).
- Define Retrieval Augmented Generation (RAG) and describe its business applications (for example, Amazon Bedrock Knowledge Bases).
- Identify AWS services that help store embeddings within vector databases (for example, Amazon OpenSearch Service, Amazon Aurora, Amazon Neptune, Amazon RDS for PostgreSQL).
- Explain the cost tradeoffs of various approaches to FM customization (for example, pre-training, fine-tuning, in-context learning, RAG, model distillation).
- Define the role of AI agents and describe AI agents' business applications.

Task Statement 3.2: Choose effective prompt engineering techniques.

Objectives:

- Define the concepts and constructs of prompt engineering (for example, context, instruction, negative prompts).
- Define techniques for prompt engineering (for example, chain-of-thought, zero-shot, single-shot, few-shot, prompt templates).
- Identify and describe the benefits and best practices for prompt engineering (for example, response quality improvement, experimentation, guardrails, discovery, specificity and concision, using multiple comments).
- Define potential risks and limitations of prompt engineering (for example, exposure, poisoning, hijacking, jailbreaking).
- Describe prompt versioning and management strategies that use Amazon Bedrock Prompt Management.

Task Statement 3.3: Describe the training and fine-tuning process for FMs.

Objectives:

- Describe the key elements of training an FM (for example, pre-training, fine-tuning, continuous pre-training, distillation).
- Define methods for fine-tuning an FM (for example, instruction tuning, adapting models for specific domains, transfer learning, continuous pre-training).
- Describe how to prepare data to fine-tune an FM (for example, data curation, governance, size, labeling, representativeness, reinforcement learning from human feedback [RLHF]).

Task Statement 3.4: Describe methods to evaluate FM performance.

Objectives:

- Determine approaches to evaluate FM performance (for example, human-in-the-loop evaluation, benchmark datasets, Amazon Bedrock Model Evaluation).

- Identify relevant metrics to assess FM performance (for example, Recall-Oriented Understudy for Gisting Evaluation [ROUGE], Bilingual Evaluation Understudy [BLEU], BERTScore, LLM-as-a-judge).
- Determine whether an FM effectively meets business objectives (for example, productivity, user engagement, task engineering).
- Identify approaches to evaluate the performance of applications built with FM (for example, RAG, agents, workflows).
- Identify business objective alignment metrics for AI applications (for example, task completion rate, user satisfaction, cost per interaction).

Content Domain 4: Guidelines for Responsible AI

Domain 4 covers guidelines for responsible AI and represents 14% of the scored content on the exam.

Tasks

- Task Statement 4.1: Explain the development of AI systems that are responsible.
- Task Statement 4.2: Recognize the importance of transparent and explainable models.

Task Statement 4.1: Explain the development of AI systems that are responsible.

Objectives:

- Identify features of responsible AI (for example, bias, fairness, inclusivity, robustness, safety, veracity).
- Explain how to use tools to identify features of responsible AI (for example, Amazon Bedrock Guardrails).
- Define responsible practices to select a model (for example, environmental considerations, sustainability).
- Identify legal risks of working with generative AI (GenAI) (for example, intellectual property infringement claims, biased model outputs, loss of customer trust, end user risk, hallucinations).
- Identify characteristics of datasets (for example, inclusivity, diversity, curated data sources, balanced datasets).

- Describe effects of bias and variance (for example, effects on demographic groups, inaccuracy, overfitting, underfitting).
- Describe tools to detect and monitor bias, trustworthiness, and truthfulness (for example, analyzing label quality, human audits, subgroup analysis, Amazon SageMaker Clarify, SageMaker Model Monitor, Amazon Augmented AI [Amazon A2I]).

Task Statement 4.2: Recognize the importance of transparent and explainable models.

Objectives:

- Describe the differences between models that are transparent and explainable and models that are not transparent and explainable.
- Describe tools to identify transparent and explainable models (for example, Amazon SageMaker Model Cards, SageMaker Clarify, Amazon Bedrock Model Evaluations, open source models, data, licensing).
- Identify tradeoffs between model safety and transparency (for example, measure interpretability and performance).
- Describe principles of human-centered design for explainable AI (for example, user-feedback mechanisms, AI decision transparency).

Content Domain 5: Security, Compliance, and Governance for AI Solutions

Domain 5 covers security, compliance, and governance for AI solutions and represents 14% of the scored content on the exam.

Tasks

Task Statement 5.1: Explain methods to secure AI systems.

Objectives:

- Identify AWS services and features to secure AI systems (for example, IAM roles, policies, and permissions; encryption; Amazon Macie; AWS PrivateLink; AWS shared responsibility model; Amazon Bedrock AgentCore Identity; Policy in AgentCore; Amazon Bedrock Guardrails).
- Describe the concept of source citation and documenting data origins (for example, data lineage, data cataloging, Amazon SageMaker Model Cards).
- Describe best practices for secure data engineering (for example, assessing data quality, implementing privacy-enhancing technologies, data access control, data integrity).
- Describe security and privacy considerations for AI systems (for example, application security, threat detection, vulnerability management, infrastructure protection, prompt injection, encryption at rest and in transit, data leakage prevention, output filtering and validation, audit trail and logging requirements for AI interactions, toxicity).
- Describe hallucination detection methods and grounding techniques to improve output accuracy (for example, Retrieval Augmented Generation [RAG] grounding, output validation, confidence scoring).

Task Statement 5.2: Recognize governance and compliance regulations for AI systems.

Objectives:

- Identify AWS services and features to assist with governance and regulation compliance (for example, AWS Config, Amazon Inspector, AWS Audit Manager, AWS Artifact, AWS CloudTrail, AWS Trusted Advisor).
- Describe data governance strategies (for example, data lifecycles, logging, residency, monitoring, observation, retention).
- Describe processes to follow governance protocols (for example, policies, review cadence, review strategies, governance frameworks such as the Generative AI Security Scoping Matrix, transparency standards, team training requirements).

In-Scope AWS Services

The following list contains AWS services and features that are in scope for the AWS Certified AI Practitioner (AIF-C01) exam. This list is non-exhaustive and is subject to change. AWS offerings appear in categories that align with the offerings' primary functions.

Topics

Analytics

- AWS Data Exchange
- Amazon EMR
- AWS Glue
- AWS Glue DataBrew
- AWS Lake Formation
- Amazon OpenSearch Service
- Amazon Quick
- Amazon Redshift

Cloud Financial Management

- AWS Budgets
- AWS Cost Explorer

Compute

- Amazon EC2

- AWS Lambda

Containers

- Amazon Elastic Container Service (Amazon ECS)
- Amazon Elastic Kubernetes Service (Amazon EKS)

Database

- Amazon Aurora
- Amazon DocumentDB (with MongoDB compatibility)
- Amazon DynamoDB
- Amazon ElastiCache
- Amazon Neptune
- Amazon RDS

Developer Tools

- Kiro
- Strands Agents
- Amazon Q

Machine Learning

- Amazon Augmented AI (Amazon A2I)
- Amazon Bedrock
- Amazon Bedrock AgentCore
- Amazon Comprehend
- Amazon Kendra
- Amazon Lex
- Amazon Nova
- Amazon Personalize

- Amazon Polly
- Amazon Rekognition
- Amazon SageMaker AI
- Amazon SageMaker JumpStart
- Amazon Textract
- Amazon Transcribe
- Amazon Translate
- AWS Transform

Management and Governance

- AWS CloudTrail
- Amazon CloudWatch
- AWS Config
- AWS Trusted Advisor
- AWS Well-Architected Tool

Networking and Content Delivery

- Amazon CloudFront
- Amazon VPC

Security, Identity, and Compliance

- AWS Artifact
- AWS Audit Manager
- AWS Identity and Access Management (IAM)
- Amazon Inspector
- AWS Key Management Service (AWS KMS)
- Amazon Macie
- AWS Secrets Manager

Storage

- Amazon S3
- Amazon S3 Glacier

Out-of-Scope AWS Services

The following list contains AWS services and features that are out of scope for the exam. This list is non-exhaustive and is subject to change. AWS offerings that are entirely unrelated to the target job roles for the exam are excluded from this list:

Topics

- Analytics
- Application Integration
- Business Applications
- Cloud Financial Management
- Compute
- Containers
- Customer Enablement
- Database
- Developer Tools
- End User Computing
- Frontend Web and Mobile
- Internet of Things (IoT)
- Machine Learning
- Management and Governance
- Media
- Migration and Transfer
- Networking and Content Delivery
- Security, Identity, and Compliance
- Storage

Analytics

- AWS Clean Rooms
- Amazon CloudSearch
- Amazon Managed Streaming for Apache Kafka (Amazon MSK)

Application Integration

- Amazon AppFlow
- Amazon MQ
- Amazon Simple Workflow Service (Amazon SWF)

Business Applications

- Amazon Chime
- Amazon Pinpoint
- Amazon Simple Email Service (Amazon SES)
- AWS Supply Chain
- AWS Wickr
- Amazon WorkMail

Cloud Financial Management

- AWS Application Cost Profiler
- AWS Billing Conductor
- AWS Marketplace

Compute

- AWS App Runner
- AWS Elastic Beanstalk
- EC2 Image Builder

- Amazon Lightsail

Containers

- Red Hat OpenShift Service on AWS (ROSA)

Customer Enablement

- AWS IQ
- AWS Managed Services (AMS)
- AWS re:Post Private
- AWS Support

Database

- Amazon Keyspaces (for Apache Cassandra)
- Amazon Quantum Ledger Database (Amazon QLDB)
- Amazon Timestream

Developer Tools

- AWS AppConfig
- AWS Application Composer
- AWS CloudShell
- Amazon CodeCatalyst
- AWS CodeStar
- AWS Fault Injection Service
- AWS X-Ray

End User Computing

- Amazon AppStream 2.0

- Amazon WorkSpaces
- Amazon WorkSpaces Thin Client
- Amazon WorkSpaces Web

Frontend Web and Mobile

- AWS Amplify
- AWS AppSync
- AWS Device Farm
- Amazon Location Service

Internet of Things (IoT)

- AWS IoT Analytics
- AWS IoT Core
- AWS IoT Device Defender
- AWS IoT Device Management
- AWS IoT Events
- AWS IoT FleetWise
- FreeRTOS
- AWS IoT Greengrass
- AWS IoT 1-Click
- AWS IoT RoboRunner
- AWS IoT SiteWise
- AWS IoT TwinMaker

Machine Learning

- AWS DeepComposer
- AWS HealthImaging
- AWS HealthOmics

- Amazon Monitron
- AWS Panorama

Management and Governance

- AWS Control Tower
- AWS Health Dashboard
- AWS Launch Wizard
- AWS License Manager
- Amazon Managed Grafana
- Amazon Managed Service for Prometheus
- AWS OpsWorks
- AWS Organizations
- AWS Proton
- AWS Resilience Hub
- AWS Resource Explorer
- AWS Resource Groups
- AWS Systems Manager Incident Manager
- AWS Service Catalog
- Service Quotas
- AWS Telco Network Builder
- AWS User Notifications

Media

- Amazon Elastic Transcoder
- AWS Elemental MediaConnect
- AWS Elemental MediaConvert
- AWS Elemental MediaLive
- AWS Elemental MediaPackage

- AWS Elemental MediaStore
- AWS Elemental MediaTailor
- Amazon Interactive Video Service (Amazon IVS)
- Amazon Nimble Studio

Migration and Transfer

- AWS Application Discovery Service
- AWS Application Migration Service
- AWS Database Migration Service (AWS DMS)
- AWS DataSync
- AWS Mainframe Modernization
- AWS Migration Hub
- AWS Snow Family
- AWS Transfer Family

Networking and Content Delivery

- AWS App Mesh
- AWS Cloud Map
- AWS Direct Connect
- AWS Global Accelerator
- AWS Private 5G
- Amazon Route 53
- Amazon Route 53 Application Recovery Controller
- Amazon VPC IP Address Manager (IPAM)

Security, Identity, and Compliance

- AWS Certificate Manager (ACM)
- AWS CloudHSM

- Amazon Cognito
- Amazon Detective
- AWS Directory Service
- AWS Firewall Manager
- Amazon GuardDuty
- AWS IAM Identity Center
- AWS Payment Cryptography
- AWS Private Certificate Authority
- AWS Resource Access Manager (AWS RAM)
- AWS Security Hub
- Amazon Security Lake
- AWS Shield
- AWS Signer
- Amazon Verified Permissions
- AWS WAF

Storage

- AWS Backup
- AWS Elastic Disaster Recovery

Revisions

AWS exam guides are periodically reviewed and updated to ensure that our certification exams test skills and AWS services and features that are relevant for the job role(s) that a certification is designed to target. Exam guide updates will be published approximately one month before updates will be reflected on your exam.

Topics

Change History

Version	Publication date
1.0	March 26, 2026
1.1	April 30, 2026

Changes to objectives

Version 1.0	Version 1.1
Objective 1.1.1: Define basic AI terms (for example, AI, ML, deep learning, neural networks, computer vision, natural language processing [NLP], model, algorithm, training and inferencing, bias, fairness, fit, large language models [LLMs]).	Objective 1.1.1: Define basic AI terms (for example, AI, ML, deep learning, neural networks, computer vision, natural language processing [NLP], model, algorithm, training and inferencing, bias, fairness, fit, large language model [LLM], generative AI [GenAI], agentic AI).
Objective 1.1.2: Describe the similarities and differences between AI, ML, GenAI, and deep learning.	Objective 1.1.2: Describe the similarities and differences between AI, ML, GenAI, deep learning, and agentic AI.
Objective 1.1.3: Describe various types of inferencing (for example, batch, real-time).	Objective 1.1.3: Describe various types of inferencing (for example, batch, real-time, asynchronous, serverless).
Objective 1.1.5: Describe supervised learning, unsupervised learning, and reinforcement learning.	Objective 1.1.5: Describe different types of AI/ML learning (for example, supervised learning, unsupervised learning, reinforcement learning methods).
Objective 1.2.4: Identify examples of real-world AI applications (for example, computer vision, NLP, speech recognition, recommend ation systems, fraud detection, forecasting).	Objective 1.2.4: Identify examples of real-world AI applications (for example, computer vision, NLP, speech recognition, recommend

Version 1.0	Version 1.1
	ation systems, fraud detection, forecasting, knowledge bases, agentic AI).
Objective 1.3.1: Describe components of an ML pipeline (for example, data collectio n, exploratory data analysis [EDA], data pre-processing, feature engineering, model training, hyperparameter tuning, evaluation, deployment, monitoring).	Objective 1.3.1: Describe and differentiate components of an AI/ML pipeline.
Objective 1.3.4: Identify relevant AWS services and features for each stage of an ML pipeline (for example, SageMaker AI, SageMaker Data Wrangler, SageMaker Feature Store, SageMaker Model Monitor).	Objective 1.3.4: Identify relevant AWS services and features for each stage of an AI/ML pipeline (for example, Amazon Bedrock, Amazon Q, Amazon Quick, Kiro, SageMaker AI).
Objective 1.3.6: Describe model performance metrics (for example, accuracy, Area Under the Curve [AUC], F1 score) and business metrics (for example, cost per user, development costs, customer feedback, return on investmen t [ROI]) to evaluate ML models.	Objective 1.3.6: Describe model performan ce metrics (for example, accuracy, precision , recall, F1 score) and business metrics (for example, cost per user, development costs, customer feedback, return on investment [ROI]) to evaluate ML models.
Objective 2.2.1: Describe the advantages of GenAI (for example, adaptability, responsiv eness, simplicity).	Objective 2.2.1: Describe the advantages of GenAI (for example, adaptability, responsiv eness, conversational capabilities, ability to generate content).
Objective 2.2.3: Identify factors to consider when selecting GenAI models (for example, model types, performance requirements, capabilities, constraints, compliance).	Objective 2.2.3: Identify factors to consider when selecting GenAI models (for example, model types, performance requirements, capabilities, constraints, compliance, cost, latency, model complexity).

Version 1.0	Version 1.1
Objective 2.2.4: Determine business value and metrics for GenAI applications (for example, cross-domain performance, efficienc y, conversion rate, average revenue per user, accuracy, customer lifetime value).	Objective 2.2.4: Determine business value and metrics for GenAI applications (for example, cross-domain performance, ROI, efficiency, conversion rate, average revenue per user, accuracy, customer lifetime value).
Objective 2.3.1: Identify AWS services and features to develop GenAI applications (for example, Amazon SageMaker JumpStart , Amazon Bedrock PartyRock, Amazon Q, Amazon Bedrock Data Automation).	Objective 2.3.1: Identify AWS services and features to develop GenAI applications (for example, Amazon Bedrock, Amazon SageMaker AI, SageMaker JumpStart, Amazon Quick, Kiro, Strands Agents, Amazon Bedrock AgentCore).
Objective 3.1.5: Explain the cost tradeoffs of various approaches to FM customization (for example, pre-training, fine-tuning, in-context learning, RAG).	Objective 3.1.5: Explain the cost tradeoffs of various approaches to FM customization (for example, pre-training, fine-tuning, in-context learning, RAG, model distillation).
Objective 3.1.6: Describe the role of agents in multi-step tasks (for example, Amazon Bedrock Agents, agentic AI, model context protocol).	Objective 3.1.6: Define the role of AI agents and describe AI agents' business applications.
Objective 3.4.1: Determine approaches to evaluate FM performance (for example, human evaluation, benchmark datasets, Amazon Bedrock Model Evaluation).	Objective 3.4.1: Determine approaches to evaluate FM performance (for example, human-in-the-loop evaluation, benchmark datasets, Amazon Bedrock Model Evaluation).
Objective 3.4.2: Identify relevant metrics to assess FM performance (for example, Recall-Oriented Understudy for Gisting Evaluation [ROUGE], Bilingual Evaluation Understudy [BLEU], BERTScore).	Objective 3.4.2: Identify relevant metrics to assess FM performance (for example, Recall-Oriented Understudy for Gisting Evaluation [ROUGE], Bilingual Evaluation Understudy [BLEU], BERTScore, LLM-as-a-judge).

Version 1.0	Version 1.1
Objective 4.2.2: Describe tools to identify transparent and explainable models (for example, Amazon SageMaker Model Cards, open source models, data, licensing).	Objective 4.2.2: Describe tools to identify transparent and explainable models (for example, Amazon SageMaker Model Cards, SageMaker Clarify, Amazon Bedrock Model Evaluations, open source models, data, licensing).
Objective 4.2.4: Describe principles of human-centered design for explainable AI.	Objective 4.2.4: Describe principles of human-centered design for explainable AI (for example, user-feedback mechanisms, AI decision transparency).
Objective 5.1.1: Identify AWS services and features to secure AI systems (for example, IAM roles, policies, and permissions; encryptio n; Amazon Macie; AWS PrivateLink; AWS shared responsibility model).	Objective 5.1.1: Identify AWS services and features to secure AI systems (for example, IAM roles, policies, and permissions; encryptio n; Amazon Macie; AWS PrivateLink; AWS shared responsibility model; Amazon Bedrock AgentCore Identity; Policy in AgentCore; Amazon Bedrock Guardrails).
Objective 5.1.4: Describe security and privacy considerations for AI systems (for example, application security, threat detection , vulnerability management, infrastructure protection, prompt injection, encryption at rest and in transit).	Objective 5.1.4: Describe security and privacy considerations for AI systems (for example, application security, threat detection , vulnerability management, infrastructure protection, prompt injection, encryption at rest and in transit, data leakage preventio n, output filtering and validation, audit trail and logging requirements for AI interactions, toxicity).

Objectives added

- Objective 1.2.6: Identify when traditional ML models versus foundation models (FMs) are appropriate for a given use case (for example, due to regulatory concerns, explainability, operational constraints).

- Objective 2.1.4: Describe the token-based pricing model and its effect on cost and performance for inference.
- Objective 2.1.5: Describe the role of context engineering in FM applications.
- Objective 2.1.6: Define foundational agentic AI concepts (for example, multi-agent system patterns for complex AI applications, Model Context Protocol [MCP] and its role in connecting agents to external systems, multi-agent communication patterns, memory management, tool usage, and workflow orchestration).
- Objective 3.2.5: Describe prompt versioning and management strategies that use Amazon Bedrock Prompt Management.
- Objective 3.4.5: Identify business objective alignment metrics for AI applications (for example, task completion rate, user satisfaction, cost per interaction).
- Objective 5.1.5: Describe hallucination detection methods and grounding techniques to improve output accuracy (for example, Retrieval Augmented Generation [RAG] grounding, output validation, confidence scoring).

Changes to in- and out-of-scope services

Services added to the in-scope list

- Amazon Aurora
- Amazon Bedrock AgentCore
- Kiro
- Strands Agents
- Amazon Q
- Amazon SageMaker JumpStart
- AWS Transform

Services removed from the in-scope list

- Amazon MemoryDB

Services removed from the out-of-scope list

- AWS DeepComposer

- Amazon FinSpace
- Amazon Honeycode
- AWS IAM Identity Center
- AWS Marketplace
- AWS Organizations
- Amazon WorkDocs

Survey

How useful was this exam guide? Let us know by taking our survey.